CRANFIELD UNIVERSITY

LINDA ANTOUN YOUSEF ALRABADY

AN ONLINE-INTEGRATED CONDITION MONITORING AND PROGNOSTICS FRAMEWORK FOR ROTATING EQUIPMENT

SCHOOL OF ENGINEERING
Department Of Power And Propulsion

This thesis is submitted in fulfilment of the requirements for the degree of PhD In Mechanical Engineering
Academic Year: 2008 - 2014

Supervisor: Professor David Mba
October 2014

ABSTRACT

Detecting abnormal operating conditions, which will lead to faults developing later, has important economic implications for industries trying to meet their performance and production goals. It is unacceptable to wait for failures that have potential safety, environmental and financial consequences. Moving from a "reactive" strategy to a "proactive" strategy can improve critical equipment reliability and availability while constraining maintenance costs, reducing production deferrals and decreasing the need for spare parts. Once the fault initiates, predicting its progression and deterioration can enable timely interventions without risk to personnel safety or to equipment integrity.

This work presents an online-integrated condition monitoring and prognostics framework that addresses the above issues holistically. The proposed framework aligns fully with ISO 17359:2011 and derives from the I-P and P-F curve. Depending upon the running state of machine with respect to its I-P and P-F curve an algorithm will do one of the following:

(1) Predict the ideal behaviour and any departure from the normal operating envelope using a combination of Evolving Clustering Method (ECM), a normalised fuzzy weighted distance and tracking signal method.

(2) Identify the cause of the departure through an automated diagnostics system using a modified version of ECM for classification.

(3) Predict the short-term progression of fault using a modified version of the Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS), called here MDENFIS and a tracking signal method.

(4) Predict the long term progression of fault (Prognostics) using a combination of Autoregressive Integrated Moving Average (ARIMA)- Empirical Mode Decomposition (EMD) for predicting the future input values and MDENFIS for predicting the long term progression of fault (output).

The proposed model was tested and compared against other models in the literature using benchmarks and field data. This work demonstrates four noticeable improvements over previous methods:

(1) Enhanced testing prediction accuracy, (2) comparable processing time if not better, (3) the ability to detect sudden changes in the process and finally (4) the ability to identify and isolate the problem source with high accuracy.

Keywords:

Condition Monitoring, Prognostics, Short Term Prediction, Long Term Prediction, Online, Automated Diagnostics, Clustering, Empirical Model Decomposition, Autoregressive Moving Average, Particle Swarm optimisation, Fuzzy Logic, Neural Network.

ACKNOWLEDGEMENTS

“I can do all things through him who strengthens me.” Philippians 4:13

I am thankful first and above all to my Lord, God almighty for his fulfilled promises and for always being there to strengthen me in the difficult times. Without him I can do nothing.

The current work is a result of assistance, guidance and support from several people. As such I would like to thank every one of them for his/her role in making this successful.

I would like to thank my supervisor Dr. Mba, who expertly guided me through all stages of my PhD. I am thankful for your advice, critical comments and support.

I would like to also thank my husband and champion Yousef Rabadi to whom I owe alot, I can’t even start counting the blessings having you in my life. My sweet heart and bundle of Joy, Helen, Mum I pray that God will give me the power to raise you and see you successful in everything you do.

I am indebted to my parents, whose value to me grows with time. Mum and Dad I hope that this will make you proud of me, you were always in mind while working hard to complete this work with all the other responsibilities I had to manage too.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ANFIS	Adaptive NeuroFuzzy Inference System
AI	Artificial Intelligence
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
BPNN	Back Propagation Neural Network
BARTFIS	Bayesian ART-based Fuzzy Inference System
BWR	Benedict-Webb-Rubin
BEP	Best Efficiency Point
BEMD	Bi-dimensional Empirical Mode Decomposition
BS-EMD	B-Spline Empirical Mode Decomposition
BIT	Built-in Test
CART	Classification And Regression Tree
CBM	Condition Based Maintenance
CM	Condition Monitoring
ECMc	Constrained Evolving Clustering Method
CHC	Convex Hull Classifier
CF	Crest Factor
DAQ	Data Acquisition
DPNN	Deep Belief Neural Network
DGA	Dissolved Gas Analysis
DWNN	Dynamic (Recurrent) Wavelet Neural Networks
DENFIS	Dynamic Evolving NeuroFuzzy Inference System
ELAM	Effective Local Approximation Method
EMD	Empirical Modes Decomposition
EEMD	Ensemble Empirical Model Decomposition
ETTF	Estimated Time to Failure
ECM	Evolving Clustering Method
EFuNN	Evolving Fuzzy Neural Networks
ESOM	Evolving Self-organising Maps
eTS	Evolving Takagi Sugeno
EWN	Evolving Wavelet Network
FMEA	Failure Modes and Effect Analysis

FMECA	Failure Modes Effects and Criticality Analysis
FAOS-PFNN	Fast and Accurate Online self-organizing scheme for Parsimonious Fuzzy Neural Networks
FFT	Fast Fourier Transformation
FFNN	Feed Forward (Static) Neural Network
FEM	Finite Element Method
FEC	Fuzzy Entropy Clustering Method
FIS	Fuzzy Inference System
FL	Fuzzy Logic
GDFNN	Generalised Dynamic Fuzzy Neural Network
GRNN	Generalized Regression Neural Network
GA	Genetic Algorithms
HMM	Hidden Markov Model
ICMF	Integrated Condition Monitoring Framework
IMFs	Intrinsic Mode Functions
LDC	Linear Discriminant Classifier
LOWESS	Locally Weighted Scatter Plot Smoothing
MTS	Mahalanobis Taguchi System
MAD	Mean Absolute Deviation
MSE	Mean Square Error
MTBF	Mean Time to Failure
MFs	Membership Functions
MDENFIS	Modified Dynamic Evolving NeuroFuzzy Inference System
MPSO	Modified Particle Swarm Optimisation Method
MA	Moving Average
MIMO	Multi-Inputs Multi-Outputs
MISO	Multi-Inputs Single Output
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MSET	Multivariate State Estimation Technique
NaN	Not a Number
NF	NeuroFuzzy Models
NDEI	Non-Dimensional Error Index
NOE	Normal Operating Envelope

OEM	Original Equipment Manufacturer
PD	Partial Discharge
PSO	Particle Swarm Optimisation
pk-pk	Peak to Peak
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
QDC	Quadratic Discriminant Class
RBF	Radial Basis Function
RNN	Recirculation Neural Network
RVM	Relevance Vector Machine
RAN	Resource-Allocating Network
RCFA	Root Cause Failure Analysis
RMS	Root Mean Square
RMSE	Root Mean Square Error
SOM	Self-Organising Maps
STFT	Short Time Fourier Transformation
SMEs	Subject Matter Experts
SVM	Supper Vector Machine
TSK	Takagi-Sugeno-Kang
WPT	Wavelet Packet Transform
WT	Wavelet Transform
0-pk	Zero to Peak
WPT-EMD	Wavelet Packet Transform- Empirical Mode Decomposition

1 INTRODUCTION

1.1 Overview

The history of Maintenance goes back even to the time of the ancient Egyptians 5000 years ago when they built an extraordinary and enduring civilisation culminating in the great pyramids. Even after this long time they are still standing there to remind us of the highly refined skills that were utilised to build them and the care that was given through the periodic inspections to every part of them. Even in more recent times, simple but effective method of using a canary as a sensor for detecting the existence of gas contaminations in the air was utilised in the coal mining industry. Since the canaries are very sensitive to the quality of air, they will lose consciousness even with low concentration of toxic or hazardous gases like carbon monoxide and methane, miners used them as gas detectors - as long as the birds are alert, the air is fresh. Otherwise a bird losing consciousness indicates immediate response such as evacuation for the whole mine due to air contamination.

In modern commercial production industry, the trend is increasing towards the need for high availability equipment which are working 24/7. This means that any failure, even if minor, is unwanted due to its production and economic impact.

Maintenance techniques involve applying one or more methods to restore equipment serviceability or to ensure equipment remains operational to the required performance. The following types of techniques may be used:

- Detection
- Inspection

- Repair/restore
- Exchange
- Re-design

The increased complexity and criticality of equipment over the years has resulted in a drastic change in the way owners used to monitor, maintain and repair their equipment.

Pre 1950's one of the main maintenance methods used was what is called "Corrective Maintenance ". This type of maintenance is carried out when a failure occurs. Other names for it include "run to failure" or "break down maintenance". There are three significant drawbacks from only utilising this maintenance method:

- Safety of workers: certain failures are catastrophic and may cause death to workers.
- Production related: due to the unexpected nature of these failures in terms of time and the lead time needed to order and replace parts which in some cases might take weeks if not months.
- Environmental related

Preventive maintenance was used in WWII and before – even the Romans had an upkeep by exchange scheme for their armies hardware (effectively planned maintenance), and Condition Monitoring was used by "wheel tappers" on Railways 150 years ago.

Preventive Maintenance is mainly a time or cycle-based method where certain inspections and PMs are scheduled at certain frequencies to be conducted. For example: bearing replacement every 2 years, oil changes every year or based on the number of working hrs or cycles, etc. Usually these tasks are based on many factors like the original equipment manufacturer's recommendations referenced in the equipment maintenance manual and sometimes based on experience. Although preventive maintenance promises to solve many of the problems caused by adapting wrong maintenance practises like corrective maintenance through reducing the number of unplanned shutdowns however it doesn't eliminate these failures completely and there are always occasions which result in the requirement for corrective maintenance. In addition, this type of maintenance can be very costly, in terms of replacing components before their end of useful life, and necessitates keeping huge inventory of components which are needed in case of failure and most important wasting resources. Preventive maintenance only applies to age-related failure modes, and has little or no effect on improving availability on other failure modes (random failure modes).

In recent times, industry has moved to applying condition based (or predictive) approach to maintenance based on trending and analysing the data from one or more parameters which indicate the presence or development of known failure modes and fault conditions. This is managed by collecting data like process parameters (temperature, pressure, flow rate, power consumption, etc.) And other health indicators like Vibration, Noise, Current signature, etc. (Jardine et al. 2006).

Analysis of this data helps in assessing the overall health of equipment and it is then possible to plan required maintenance based on the severity of the detected fault (Vachtsevanos et al. 2007), equipment condition and its criticality to the production line. Condition based maintenance can be effective for most failure modes, including age related and random.

Researchers in conjunction with industry have started to focus on this type of maintenance (Predictive Maintenance) to enhance its deliverables through utilising different techniques of measurement and signal processing, fusing data from different sources to reduce the uncertainties related to human and instrumentations errors,…etc. In modern days there has been a trend towards automating the whole process. That leads to the desire of the engineering management and planners to have information such as the estimated useful life of their equipment and the ability of the equipment to perform the required tasks until the next planned shutdown successfully. This involves diagnostics and prognostics and results in optimisation of maintenance planning and cost.

The rapid growth of Maintenance Methods from Corrective Maintenance (Run until shutdown) through Preventive (Time based) to Predictive Maintenance (Condition based) decreased the trends of unplanned shutdown events and production deferrals.

This avoids supporting the cost of huge amount of spare parts inventory and enables proper planning for spare parts ordering, taking into consideration the machine condition and the lead time of delivery of these replacement parts. Most important it improves the safety of personnel as a result of avoiding unmonitored catastrophic failures of machinery.

The advancements in the areas of data collection, signal processing, transducers technologies and continuous improvement of diagnostic and prognostic models have enabled the development of the CBM approach.

1.2 Problem Statement

Rotating equipment are considered the backbone and major component of any oil and gas plant. Like any other type of equipment, these equipment are designed to run for many years trouble-free, however in reality, there is a difference between the equipment designed Mean Time to Failure (MTBF) and the actual MTBF due to many factors including but not limited to: poor design, wrong assembly and installation, running the equipment at extreme operating and environmental conditions, etc. 80% of rotating equipment failures are random (meaning that there is no obvious relation between how long the equipment has been in service and the likelihood of failure happening) as compared to 20% aged related failures however rotating equipment have distinctive characteristics that can be monitored and analysed to give an indication about their overall health. Moving from a “reactive” maintenance strategy to a “proactive” maintenance strategy can improve critical equipment reliability and availability while constraining maintenance costs, reducing production deferrals and decreasing the need for spare parts. In support of a proactive maintenance strategy, condition monitoring has gained a lot of interest from researchers and industries. Many developments and novelty approaches were presented, applied and proven to have excellent outcomes. However majority of these approaches rely heavily on human expertise, as people move on, the knowledge moves with them creating sustainability issues when applying these approaches adding to that the likelihood of human errors. As a result, an increasing trend towards automating the process was seen in recent years. Detecting abnormal operating conditions, which will lead to faults developing later, has important economic implications for industries trying to meet their performance and production goals. It is unacceptable to wait for failures that have potential safety, environmental and financial consequences to happen. Once the fault initiates, predicting its progression and deterioration can enable timely interventions without any risk to personnel safety or to equipment integrity.

This prediction can help in two sides: check whether the equipment will survive until the next planned shutdown and check the remaining useful life of the equipment also called prognostics. The short and long term predictions are normally calculated using the present and past equipment health taking into consideration any variations in the operating conditions in the past, present and future. During the life cycle of the equipment some routine maintenance tasks may help improving the overall health of the equipment like for example applying lubrication to bearings regularly while some others can potentially have an adverse effect on the overall health of the equipment again using the previous example if the routine task is done in the wrong way by over greasing a bearing, this put the bearing under too much stress and eventually cause a bearing failure.

It’s important to be able to detect these sudden changes in the process and act accordingly. To enable an informed maintenance planning for the work to be done rather than by trial and error which will definitely extend the downtime duration, the type of failure mode needs to be identified either through an automated or manual process. Subject Matter Experts are normally involved either ways.

Different models were applied for diagnostics and prediction purposes ranging from statistical and reliability models, physical based models, data driven models and hybrids of more than one model. Majority of the work done assumed that the data is available upfront to train, validate and test the proposed models when in reality this isn't the case for example: with a newly installed equipment there is no historical data available. As such an online model is needed that updates itself continuously with any newly acquired data/knowledge. This is a very important feature to apply the model across both new and old equipment, to avoid limitations in previous models caused by data availability which cover a wide range of normal and faulty operating conditions but most important to detect transient changes resulting from instrumentation errors, actual maintenance/repair events, etc. This work presents an online-integrated condition monitoring and prognostics framework that addresses the above issues holistically.

1.3 Scope of Research

Even though the proposed approach is applicable to any type of equipment (rotating, reciprocating and static), this work only looks at rotating equipment and implements the proposed model on a fouled centrifugal compressor.

Within the condition based maintenance framework, this work covers three main parts:

- Monitoring the departure of the equipment health indicators outside their normal operating envelope.
- Isolate and identify the fault developed through an automated diagnostics system

- Predict the future progression of the fault both short term and long term (prognostics)

There are some commonalities between the algorithms used for the above three tasks, due to the flexibility of the proposed model and its wide applications in the fields of prediction and classification.

The proposed model is a hybrid between two data driven techniques, namely: Neural Networks and Fuzzy Logic. This is supported by a global optimisation method (Modified Particle Swarm Optimisation) for fine tuning the solution obtained by the online grid partitioning method (Evolving Clustering Method). For both short and long term predictions, there are two scenarios. The first scenario assumes that the future inputs to the model are available and as such the model can be used for single step ahead and multi-step ahead predictions. The second scenario assumes that the future inputs to the model aren't available, in real field applications this is normally the case. In the second scenario, a method is required to predict the future values of the inputs based on their past and present values, for this purpose a hybrid between the empirical mode decomposition and autoregressive moving average methods is proposed. The ECM for classification modified algorithm is proposed for the automated diagnostics task.

Even though the proposed approach is applicable to any type of equipment (rotating, reciprocating and static), this work only looks at rotating equipment and implements the proposed model on a fouled centrifugal compressor.

Within the condition based maintenance framework, this work covers three main parts:

- Monitoring the departure of the equipment health indicators outside their normal operating envelope.
- Isolate and identify the fault developed through an automated diagnostics system
- Predict the future progression of the fault both short term and long term (prognostics)

There are some commonalities between the algorithms used for the above three tasks, due to the flexibility of the proposed model and its wide applications in the fields of prediction and classification.

The proposed model is a hybrid between two data driven techniques, namely: Neural Networks and Fuzzy Logic. This is supported by a global optimisation method (Modified Particle Swarm Optimisation) for fine tuning the solution obtained by the online grid partitioning method (Evolving Clustering Method). For both short and long term predictions, there are two scenarios. The first scenario assumes that the future inputs to the model are available and as such the model can be used for single step ahead and multi-step ahead predictions. The second scenario assumes that the future inputs to the model aren't available, in real field applications this is normally the case. In the second scenario, a method is required to predict the future values of the inputs based on their past and present values, for this purpose a hybrid between the empirical mode decomposition and autoregressive moving average methods is proposed. The ECM for classification modified algorithm is proposed for the automated diagnostics task.

1.4 Research Hypothesis

The research hypothesis are as follow:

1. The following modifications in the original DENFIS model can improve the model prediction accuracy: Using all generated rules, using exponential MFs instead of rectangular MFs. The null hypothesis is that the new proposed approach will make no difference or actually adversely affect the accuracy of the model.
2. A modified global optimisation method (MPSO) integrated with the proposed model can improve both the prediction accuracy and processing time of the model. The null hypothesis is that the new approach will make no difference.
3. Modifying the original model to learn/add new rules while working online can improve the models' capabilities in detecting sudden changes in the process. The null hypothesis is that this modification will have no effect and the model won't be able to detect sudden changes in the process.
4. Modifying the original model during the classification testing phase can improve the classification accuracy of the model. The null hypothesis is that the new approach will make no difference.

1.5 Originality and Contribution

This work presents a novel approach in the following areas:

1. Proposing a hybrid of ECM and normalised weighted fuzzy distance for the first time to monitor where the machine is running with respect to its normal operating envelope (NOE) and then alert the user if departure outside this zone occurs by using a tracking signal. The combination of

ECM and weighted fuzzy distance is in its own a new model and also the usage of this model for predicting the normal equipment behaviour and compare that with where it’s actually running is a second contribution here. The process is proposed to be monitored using a tracking signal which has the capability of identifying any departure beyond the NOE.

2. Modifying the ECM for classification algorithm to enable online updating of the clusters during the testing stage. This feature enhanced the classification accuracy of 150 cases of the Iris flower from 98.6% to 100%.
3. Enhancing the prediction accuracy of the Dynamic Evolving NeuroFuzzy Inference model by introducing the following changes to the model: using exponential membership functions instead of the originally proposed rectangular membership functions, integrating the model with a modified version of particle swarm optimisation to optimise the position of the clusters centres and width of the membership functions, and finally adding an online updating feature during the testing stage online to enhance the models capability of detecting sudden changes in the process.
4. Introducing a hybrid EMD-ARIMA model to enable predicting the future values of the inputs which in turn enables prediction of multiple steps ahead (long term prediction) of the health indicators. This modification was successfully tested on fouling progression rate and the model predicted successfully 250 days steps ahead.

1.6 Thesis Structure

This thesis consists of 6 main chapters. Chapter 1 gives an introduction to the research, research motivation, scope of research and main contributions. A summary of an extensive literature survey conducted in the field of condition monitoring and prognostics is presented in Chapter 2. Chapter 3 covers the enabling technologies and theories, the proposed model is also described in details in this chapter. The proposed model was first tested on three benchmark datasets through 5 case studies to test a number of hypothesis for improvement in the prediction accuracy, processing time and classification accuracy in Chapter 4. This is then followed by validating the proposed model using field data for fouled compressor and electric transformers in Chapter 5.

Conclusions and recommendations are presented in Chapter 6. This chapter summarises the conclusions in light of the benchmarks and field datasets simulations, list the main contributions and proposes some future work.

2 LITERATURE SURVEY

2.1 An Introduction to Condition Based Maintenance (CBM)

CBM is a maintenance strategy whereby equipment is maintained according to its condition, rather than on an elapsed time or running hour's basis. This strategy involves periodically analysing the equipment condition monitoring information like vibration, oil analysis, thermography, process data, etc. to assess its overall condition and to only carryout maintenance when required. The two main pillars of condition based maintenance strategy are diagnostics and prognostics. Diagnostics involves identifying the root cause of a problem whereby the problem has already occurred and Prognostics involves predicting the future health of the equipment either before or after a problem occurred (Jardine et al. 2006).

The implementation of any CBM strategy consists of four different phases:

1. Data collection: Data are normally acquired using online or offline condition monitoring systems. The two basic types of data are events and condition monitoring data (Jardine et al. 2006). Events data include equipment historical maintenance records and reliability information, and condition monitoring data include vibration, performance and process data, current signature analysis, oil analysis, etc. To ensure a successful CBM strategy both event and condition monitoring data should be used as they both have the same level of importance. The decision into which technology to use and at what measurement frequency is based on the failure modes likely to happen, their estimated time to failure (ETTF) and sensitivity of the condition monitoring parameters in detecting these failure modes.

2. Data analysis: Data collected are normally in raw format and might need cleaning, processing and potentially data reduction before any informed decision can be made based on this data. Cleaning involves removing wrongly assigned failure modes to certain events data, removing NaNs (Not a Number Values) and bad data. Advanced statistical and signal processing techniques can also be used to extract useful information from the data that are otherwise hidden within.

3. Decision Making: Once the data is processed and analysed the next step is to make a decision into the overall health of the equipment. This might involve an intrusive or nonintrusive actions. For instance, if the data indicates a machine running outside its normal operating envelope this could be as simple as recommending a change in its operating routines, whereas a late stage of fault development might involve a downtime to repair/replace the component involved.

4. Feedback: Once the maintenance action/change in the operating routines is taken it is vital to feedback any findings which will aid in the continuous improvement cycle of the CBM strategy.

2.1.1 Diagnostics

The foundation of any CBM strategy is robust and reliable fault diagnostic capabilities. Diagnostics algorithms are designed to assess equipment performance, monitor deterioration rates, and identify any impending failures based on certain parameters variations (Vachtsevanos et al. 2007). Ideally, such algorithms has the capability of identifying even the sub-equipment component that is failing and its failure mechanism. Over the last 50 years, an accelerating rate of publications in this area was noticed, recognising the importance of this pillar to any CBM strategy and trying to enhance its capabilities.

Some of the earliest fault diagnostic capabilities developed in the form of built-in test (BIT) for early generation aircrafts (Vachtsevanos et al. 2007). Continuous developments in the computer power and data storage capabilities and technology in general had a major influence on the accelerating trend of development in systems diagnostics capabilities as new information are made available. It is now possible to identify at an early stage the presence of incipient faults early enough to take action which could potentially reverse the failure progression process. Such capabilities enabled maintenance personnel to avoid potentially catastrophic failures and enhance the equipment reliability and availability.

2.1.2 Prognostics

Prognostics derives from the Greek word **Prognostikos** and means fore-knowing or fore-seeing, originally used in the Medical field by doctors to predict the chances that a patient will recover from a particular disease or that the disease will return back, by using statistical analysis methods based on groups of people whom suffer from same symptoms as that of this patient such that a proper treatment plan can be prepared. Due to the criticality and benefits of such kind of analysis for the Medical field the development of prognostics in the Medical field is more mature compared to other Industrial related applications (Absolute Astronomy 2009).

BS ISO 13381-1:2004 defines prognostics as: "*the estimation of time to failure and risk for one or more existing and future failure modes*" (ISO 13381-1:2004 2004). This definition introduces the need for knowing the possible failure modes associated with the system or component under study through conducting Failure Modes and Effect Analysis (FMEA) study as outlined in IEC 60812 (IEC 60812 2006) or Failure Modes Effects and Criticality Analysis (FMECA) study as outlined in BS 5760-5 (BS 5760-5:1991 1991). The casual tree relationships between past, present and future failure modes is also described in this standard, differentiating between primary, secondary and tertiary failure modes. The secondary failure mode is the failure mode initiated due to an existing primary failure mode which influences this initiation and same for the tertiary failure modes. The physics of failure mechanisms forms one of the elementary parts of any successful Diagnostics/Prognostics framework (ISO 13381-1:2004 2004). Reference (Vachtsevanos et al. 2007) described that as the "*cornerstone*". By conducting a FMECA study one can relate the failures/defects to their root cause i.e. failure mechanism.

Prognostics studies have been applied to different fields and different mechanical, electrical and electronic components and systems. Applications like: Rotating and reciprocating machinery from civil and military applications, batteries, circuit boards, specific components (bearings, gearboxes, etc.).

To maximise the benefits of a truly CBM strategy, additional capabilities, beyond the realm of diagnostics are required. Prognostics capabilities are designed to provide maintenance personnel with insight into the future health of a monitored equipment. Figure 2-1 illustrates the life cycle of an equipment from as new condition to a failed state (Vachtsevanos et al. 2007).

At the start of the equipment life cycle, the equipment is considered new, this could also be following completion of a work order and commissioning the equipment, successful transition from the infant mortality period, the equipment will continue in good working condition. After some time, an incipient fault condition develops in one of the equipment components. As time progresses, the severity of the fault increases until the component eventually fails, this is normally called the PF interval which is the time interval between the initiation of a potential fault in the component/equipment and the time component/equipment reaches its functional failure. If the equipment is allowed to continue operating beyond this point secondary and tertiary damage to other components will occur.

Diagnostics is normally done anytime between the very early incipient fault point and the sub-component functional failure depending on the sensitivity of the diagnostics system and/or competency of the analyst. However, prognostics can be done even prior to the initiation of the fault by detecting any early departure from the normal operating envelope. Prognostics has the potential to deliver major benefits as part of the CBM strategy more than any other maintenance approach by increasing the availability of the equipment and reducing the overall maintenance cost.

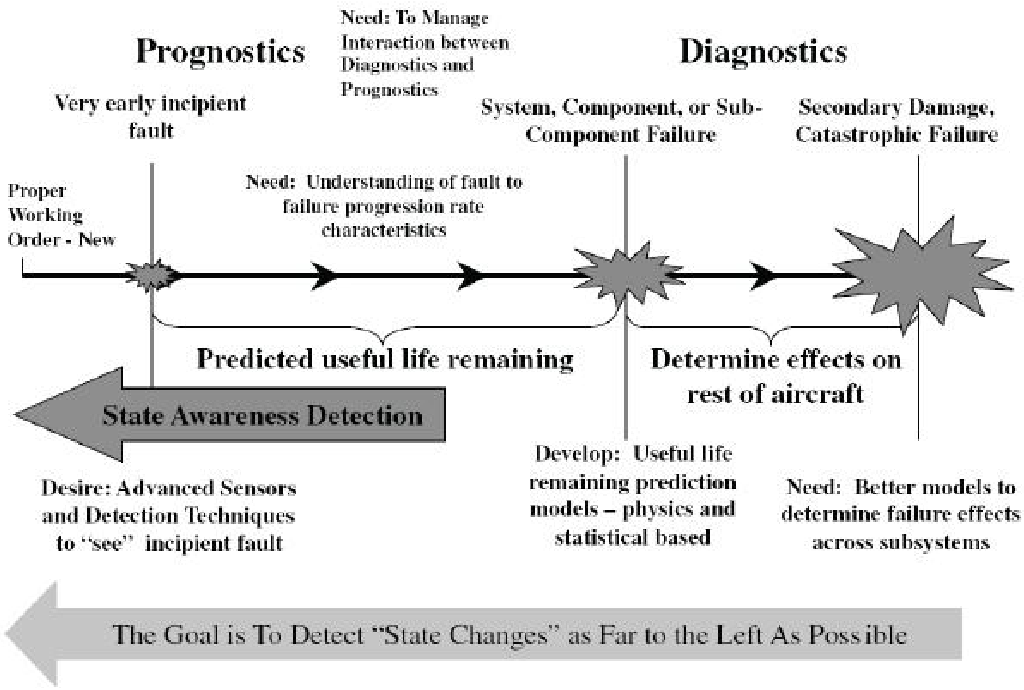


Figure 2-1 Failure progression timeline (Vachtsevanos et al. 2007)

There are different classifications for the diagnostics and prognostics methods in the literature, some are more general than the others. However, the three main approaches are (Heng et al. 2009), (Jardine et al. 2006):

- **Model-based approaches**: These approaches use physical principles and system identification methods in describing the failure initiation and growth. Due to the nature of these methods, the amount of data required with these models are minimal when compared with other approaches. However they are computationally expensive, fault specific, and not practical to be implemented on a larger scale.
- **Data-driven approaches**: These approaches depend on CM and other operational data collected to build the models. The accuracy of the output is very much linked to the amount and quality of collected data. Statistical Methods, Artificial Intelligence (AI), etc. belong to this approach.
- **Hybrid approaches**: These approaches combine either data driven and model based approaches and/or more than one data driven approach. The

objective is to improve the capabilities of the original models by combining more than one of them together to take advantage of the benefits of each of them. This approach has been adopted by the author and used through this thesis.

2.2 Condition Based Maintenance Framework

Figure 2-2 presents a CBM framework as proposed by ISO 17359:2011 (BS ISO 17359:2011 2011) where each stage performs a unique function to ensure the process of CBM can be successfully implemented.

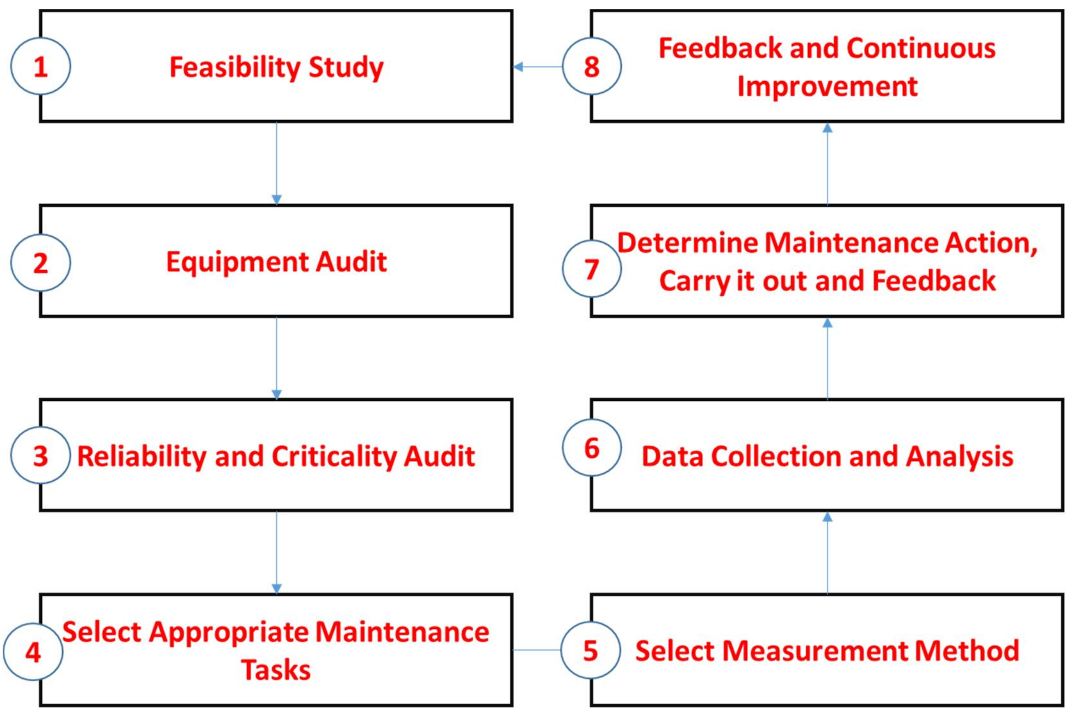


Figure 2-2 ISO 17359:2011 CBM Framework (BS ISO 17359:2011 2011)

Steps 6 & 7 of the above CBM framework will be covered in more details below.

2.2.1 Data Collection and Validation

2.2.1.1 Data Collection

The process of acquiring various equipment parameters and storing them in a central database is called Data Collection (Jardine et al. 2006). Data is normally acquired from sensors placed on the equipment either permanently or temporarily to monitor certain aspects of the mechanical and performance behaviour of the equipment. An example offline data collection and validation process is shown in Figure 2-3, a vibration sensor monitoring the relative shaft vibration at a centrifugal compressor drive end bearing side is shown, the data is transferred through the cable to a data collection portable where data is stored temporarily until they get transferred to a desktop application for further validation and analysis.

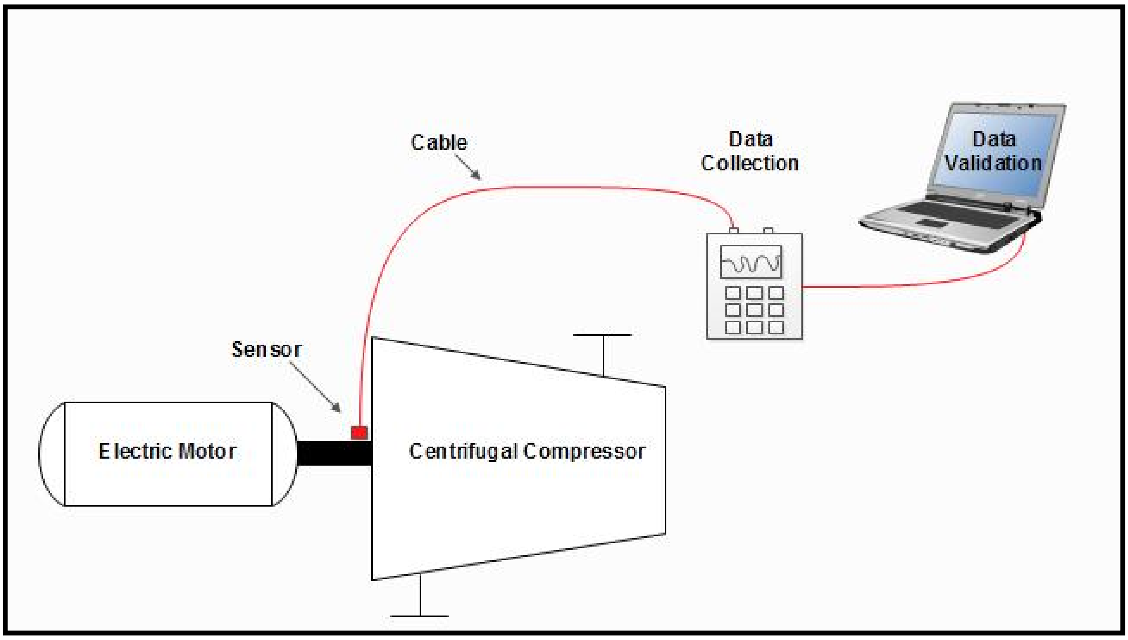


Figure 2-3 Example Offline Data Collection and Validation

This data usually falls into two categories: condition monitoring (CM) and event data (Jardine et al. 2006). Examples of CM data are vibration, oil analysis, pressure, temperature, flow rate, valve opening position, etc. Examples of event data are information obtained from maintenance overhauls, routine preventive maintenance, visual inspection, etc. CM data can be classified based on the type of data acquired into static data like pressure and temperature readings and dynamic data like vibration and noise time waveforms. Static data are normally trended over time and signs of departure from the normal operating envelope are normally captured through some predefined alarm limits either based on original equipment manufacturer's recommendations, internationally accepted standards and/or based on subject matter experts experience. Dynamic data are normally collected at a specific time, this data are usually stored in a raw form which normally require further processing, like for example transformation into the frequency domain, extracting some statistical indicators like the mean, maximum and minimum values, etc.

Data collection can be done using either online or offline monitoring system. The two main differences are related to whether the data collection process is automatic or manual and whether data is collected from permanently installed or temporarily installed sensors.

In an online monitoring system, data is collected from permanently installed sensors on the equipment at various locations, ex. Vibration, bearing metal temperature, suction pressure , discharge pressure, suction temperature, and discharge temperature, fluid (or gas) flow rate, valve opening position, filter differential pressure, lube oil temperature and pressure, etc.. This data is collected automatically at a preconfigured sampling rate ranging from real-time continuous to periodic data collection. The preconfigured sampling rate is linked to time (ex. every 10 min, hourly, daily, etc.) or an event (ex. start-up/shutdown and also if an alarm limit is exceeded). Usually critical equipment like power generation and gas compression units are equipped with online monitoring systems.

In an offline monitoring system, data is collected from permanently or temporarily installed sensors on the equipment at various locations. This data is collected manually and periodically. The decision to monitor a piece of equipment with an offline monitoring system is based on a feasibility and cost justification study, taking into consideration criticality of equipment, the failure modes likely to occur and their lead time to failure but most importantly, the consequence of failure. Sometimes it isn't even required to sample certain parameter continuously. Visual

Inspection and event data collection belongs to the offline monitoring category even though it’s not intrusive meaning that no sensors are required to be installed; it relies on the human senses. Infact this is the most commonly used offline monitoring technique in industry for all types and criticality of equipment. For all offline monitoring systems the measurement locations should be clearly identified and marked to ensure measurement repeatability

Data acquisition in its own will provide the raw data without any additional processing, however to enable a proper decision making relying on the diagnostics and prognostics outcomes, the data will almost always require further processing as raw data can include outliers, noise, nonstationary behaviour and others (Jablonski and Barszcz 2013). Using unclean data will affect the performance and quality of the overall process.

2.2.1.2 Data Validation

Whether data is collected using an online or offline monitoring system, it is vital to ensure that this data is validated before any informed decision can be made. Problems like electric interference, sudden changes in the process, sensor field wirings integrity, wrongly mounted sensors, etc. are some examples of issues that produce invalid readings (Jablonski and Barszcz 2013). Data validation in condition monitoring is the process of checking for errors in the collected data and correcting them (by replacing them, eliminate them from the analysis or mark them as invalid) to ensure high quality information are being fed to the following steps of the condition monitoring program. The objective is to only pass on complete, correct and consistent data. Data Validation is used in so many different fields (for example electrical engineering, chemical and petroleum engineering, computer science and information technologies, mechanical engineering, etc.) with various proposed methods that are applicable to some or all fields. A comprehensive literature survey into the topic is covered in (Sun, S. et al. 2011) with bias towards water systems. This survey classified data validation in 4 different ways as shown in Table 2-1:

Table 2-1 Data Validation Classifications

Class 1	**Class 2**	**Class 3**	**Class 4**
Manual	Offline	Simple tests	Single dimensional
Automatic	Online	Physical/mathematical tests	Multi-dimensional

The first two classifications (Class 1 & 2) are linked to the data collection method whether this is online or offline. The need for automatic and online data validation methods has increased over the past decade with the increasing and dynamic developments nature of technologies and moving towards real time data collection methods where processing time is crucial. Within the condition monitoring umbrella, critical rotating equipment are normally equipped with online monitoring system to collect several condition monitoring indicators including vibration, bearing metal temperature, inlet and outlet pressures and temperatures, flow rate, etc. This is normally done at high sampling rate approaching real-time data collection.

Class 3 is based on the type of test used for data validation ranging from simple tests as simple as visual examination of the data in a trend plot or using some graphical applications and as complex as using certain mathematical tests like principle component analysis and artificial neural networks.

Class 4 is based on the type of data, whether this is Univariate (single dimensional) or multivariate (multi-dimensional). In Data validation for multi-dimensional dataset the interaction between the different variables is normally studied and any abnormal behaviour is flagged. Data validation isn't going to be described in more details in the section as it isn't the main topic of the author's research however a comprehensive review is provided in Table 2-2 which include commonly used methods of data validation in the literature, their classification based on Classification III in Table 2-2, errors and example causes.

Table 2-2 Data Validation Methods, Classification Type, Error and Example Cause

Method	Classification Type	Error	Example Cause	References
Status Check	Simple test	Working Staus	Cabling shorts, open circuits, Power supply failure	na
Calibration Due Date Check	Simple test	Calibration Status	Out of Date Calibration for an accelerometer	na
Drift detection by exponentially weighted moving averages	physical or mathematical tests	Drifts	slowly evolving, gradual faults	(Ross et al. 2012), (Millioz et al. 2011)
Signal's gradient test	Simple test	Drifts	Loose connections, sensor saturation, sudden change of load	(Jablonski and Barszcz 2013), (Pouliezos and Stavrakankis 1994)
Constant Value Detection	Simple test	Flatlined Trend	Data Loss	(Jablonski and Barszcz 2013)
Presence Check	Simple test	Missing Data	Intermittent data feed disconnection	(D 3.1.1)
Limit Check (Physical Range, Local realistic range and tolerance band)	Simple test	Outliers	Incorrect probe sensitivity used, sensor settings, unit conversions	(Kim et al. 1992),(Gryllias and Antoniadis 2012)
Extreme Value Check Using Statistics	physical or mathematical tests	Outliers	Loose connections, sensor saturation, sudden change of load	(D 3.1.1)
Multivariate statistical test using PCA or kernal PCA	physical or mathematical tests	Outliers	air sample contamination	(Gribok et al. 2000), (Brown et al. 2010),(Dunia et al. 1996)
Data Mining Technology	physical or mathematical tests	Outliers and other quality issues	Various sensors issues	(D 3.1.1)
Gross Error Detection	physical or mathematical tests	Outliers	zero error (imperfect calibration of sensors), environmental changes, imperfect measurement process	(Qin and Li 1999)
Uncertainity Considerations	physical or mathematical tests	Uncertainity	Measurement uncertainities	(D 3.1.1)
Material Redundancy Detection	Simple test	lack of redundancy	damaged sensor or gradual performance deterioration	(D 3.1.1)
Spatial consistency Method	physical or mathematical tests	Uncorrelated	damaged sensor or gradual performance deterioration	(Lee 1994)
Analytical redundancy Method	physical or mathematical tests	Uncorrelated	damaged sensor or gradual performance deterioration	(Chew and Wilsky 1984)

2.2.2 Data Processing and Feature Extraction

This process involves three different phases: feature extraction, data cleaning and feature selection. Data cleaning can be done prior to the features extraction in the collected raw data domain or after extracting the features. It is important though to ensure that only clean data is passed on to the next stage which involves fault classification and diagnostics. These three phases are introduced one by one below.

2.2.2.1 Feature Extraction

In most cases, raw data cannot be used directly for diagnostic and prognostic purposes. One needs a way of obtaining useful information from the raw data; this is known as feature extraction. The useful information extracted is often referred to as the condition indicators or features; they reflect the health status of the equipment or component. Around 99% of all failures are preceded by changes in one or more health indicators (Geitner and Bloch 2012). Feature extraction is a widely studied problem for which numerous models, algorithms and tools have been developed. These are reviewed in terms of their domain here. The collected raw time waveform data can be directly analysed in the time domain by extracting useful characteristic information (features) from the signal using statistical methods. The following are examples of some extracted features: peak amplitude or zero to peak (0-pk), peak to peak (pk-pk), root mean square (RMS), Crest Factor (CF), variance and standard deviation, other higher statistical orders, etc. The definition of these extracted features is given in (Sun et al. 2004). CF is calculated using the ratio between the 0-pk and RMS values of the time waveform, it gives a good representation about the presence of impacts in the time waveform, this is especially of great importance as a health indicator for rolling element bearings. Kurtosis is based on a comparison between the 4th statistical moment and the second statistical moment squared as shown in equation 2-1. It gives an assessment about how the data is distributed around the mean value.

$$\text{Kurtosis} = \frac{\frac{1}{N}\Sigma(x_i^4)}{\left(\frac{1}{N}\Sigma(x_i^2)\right)^2} \tag{2-1}$$

In (Samanta and Nataraj 2008) the kurtosis feature was used as a condition indicator for the prognosis of gearbox condition subject to gear pitting wear. Time domain features can also be extracted using more complex statistical models. In (Samuel and Pines 2005) a comparison was made between several time domain features and their sensitivity in detecting faults of helicopter gearboxes. In (Poyhonen et al. 2004) an autoregressive model was built based on vibration data collected from a motor, the parameters of the model were used condition indicators (features) for fault identification. Other parametric time series applications can be found in (Mechefske and Mathew 1992) and (Baillie and Mathew 1996).

A time waveform is a mixture of one or more frequencies which make the signal too complicated to be analysed in the time domain. Transforming this data into a different domain might be a better option in this case. Every fault has a characteristic frequency that identifies it. Transforming the data from the time to the frequency domain will highlight all the frequencies contributing to the complex time waveform shape. Those frequencies can then be compared to the main faulty frequencies for fault identification purposes. The most popular time to frequency domain transformation technique is the Fast Fourier Transformation (FFT), however this method only applies to stationary time waveforms (constant mean and variance) else signal processing errors might be generated. Other widely used techniques are: enveloping (Stack et al. 2002), power spectral density, etc.

With some of the developed faults the stationary assumption of the time waveform isn't any more valid hence a third hybrid domain of the two was proposed in the literature to address this issue namely the Time-Frequency domain analysis. The most typical method of this kind is wavelet transform (WT).

WT has developed rapidly in the recent decades and has been widely used for feature extraction (Mori et al. 1996), (Yang et al. 1999), (Staszewski et al. 1999) and (Cheung et al. 1993). A continuous WT is described by equation 2-2 below:

$$W(a,b) = \frac{1}{a}\int_{-\infty}^{+\infty} x(t)\Psi^*(\frac{t-b}{a})dt. \tag{2-2}$$

Where:

$x(t)$ is the time waveform, a is the scale parameter, b is the time parameter and

Ψ is the WT function

Widely used WT techniques include continuous WT, discrete WT (Mori et al. 1996), (Wu and Hsu 2009), (Wu and Liu 2008) and (Wu and Kuo 2009), and wavelet packet transform (WPT) (He and Shi 2002). (He and Shi 2002) Applied a Time-Frequency domain method (WPT) to extract faulty valves features from the vibration data collected at a number of reciprocating pumps. (Saravanan and Ramachandran 2009) Extracted certain features using various wavelets for bevel gearbox failure modes classification. (Wang et al. 2010) Developed a continuous wavelet transform based approach for fault detection under variant loads. Short Time Fourier Transformation (STFT) was develop to address the non- stationary signal issue by introducing a sliding FFT window over the signal, assuming that within this window the part of the signal analysed is stationary (Gao and Yan 2006).

In (Forrester 1989a), (Forrester 1989b) and (Forrester 1989c) Wigner Ville Distribution was applied to vibration data from transmission gears for faults (ex. cracked tooth and pitting) detection.

Empirical Mode Decomposition (EMD) as a novelty time-frequency domain method was first introduced in (Huang 1999). As the name reflects, the signal is decomposed into a number of signals and a residual signal. The decomposed signals are called intrinsic mode functions (IMFs). The major IMFs are normally used for feature extraction. EMD and AR were used in (Cheng and Yang 2006) for extracting features from vibration data of roller element bearings for diagnostics purposes.

2.2.2.2 Data Cleaning

In certain cases, the extracted features using the time, frequency and time-frequency domain method are not 100% reliable and need to be further processed in order to identify and eliminate any unwanted outliers. This is known as data cleaning. This stage involves two successive actions, detecting and eliminating outliers. The key problem in data cleaning is how to accurately detect existing outliers. Classical methods of detecting outliers adopt a statistical way where some distance measure is used to find out the similarity between a particular data point and the rest of the data. The Mahalanobis distance described by equation 2-3 is an example of distance measures used for outliers detection in the multidimensional space:

$$\mathrm{MD}_i = \sqrt{(x_i - \mu)S^{-1}(x_i - \mu)^{\mathrm{T}}}, \quad \textbf{(2-3)}$$

where x is a random vector, representing attributes or features in N dimensional space, the mean value is μ, and the covariance matrix is S . The MD considers the correlation between different features.

(Rousseeuw and Zomeren 1990) Proposed to compute a distance measure derived from the covariance of data points and their robust locations estimates. (Fung 1999) Proposed a method of detecting outliers based on the S-estimation robust method. (Shieh and Hung 2009) Proposed a method for detecting outliers in microarray data using Principal Component Analysis (PCA) and Mahalanobis Distance. Other methods like: Low Pass Filtering, Time Synchronous Averaging (Coats et al. 2009) and Adaptive Noise Cancellation (Chhikara and Singh 2012) were used in the literature for data cleaning.

2.2.2.3 Features Selection

Generally speaking, feature selection for classification and diagnostics purposes is different from that for prediction and prognostics purposes. For prognostics, researchers focus mainly on Univariate time series prediction where only one feature is involved, so suitable features can be directly selected by visualizing their trends, e.g. if the values for a feature monotonically increase as time passes, this feature could be a good representative of the overall health of the equipment. For diagnostics, the rationale behind feature selection is more complex for classification problems where the dimension of features space is usually greater than one. For a certain problem, many recommended features may be extracted according to the reported literature; however, these features may not all be useful since practical cases vary with specific assumptions, constraints, etc. For cases where all features are required and the number of features is small, one can simply use all of them, but if the dimension of feature space is enormous, there will be a huge computational burden for the forthcoming stage, classification. It is, however, more common for some features to be redundant and/or irrelevant. The redundant features contribute nothing to the classification results but the feature space size, and any irrelevant features actually adversely affect the classification results; hence, such features should be selected and eliminated from the feature space.

Selection methods are generally classified into filters and wrappers. The main difference between the two is that with filters, the feature selection process is done in isolation of the learning algorithm used whereas with wrappers the leaning algorithm is included as part of the selection process (Kohavi 1995) and (Kohavi and John 1996), wrappers normally give better accuracy however on the expense of computational time. Principal Component Analysis (Malhi and Gao 2004), ranking variables based on their correlation coefficient values (Thiry et al. 2004), Support Vector Machines (Weston et al. 2001), (Chapelle et al. 2001) and (Yang et al. 2014) and decision tree algorithms (Kohavi and John 1996) are some examples of feature selection methods.

2.2.3 Fault Classification (Diagnostics)

Diagnostics is defined as the identification and isolation of a fault condition. Many approaches were published in the literature for diagnostics and classification problems. Example review papers for diagnostics approaches can be found in (Korbicz et al. 2004), (Jardine et al. 2006) and (Vachtsevanos et al. 2007). Two main steps are normally covered within this area: 1. Identify the overall health status, i.e. machine in alarm or not and 2. Isolate the fault type and location. Both will be described in more details below.

2.2.3.1 Alarms and Notifications

Alarms represent a departure of equipment condition outside its normal operating envelope. This can be in the form of absolute alarms, residuals, step change, narrow band alarms and statistical alarms. Table 2-3 shows a list of rotating and reciprocating equipment covered by ISO 10816 and ISO 7919 specifically for absolute and relative shaft vibration severity limits. These limits are normally used as a guide if the Original Equipment Manufacturer (OEM) limits aren't defined or specified.

Setting up residual alarms based on the difference between actual and ideal behavior is also used to detect at an early stage deviations from the normal operating envelope in which case an action can be taken which could potentially reverse the failure progression process and save the equipment from a catastrophic failure (Wegerich 2005). Figure 2-4 describes the fault detection and identification process as proposed in (Wegerich 2005). After selecting the appropriate features, extracting the data and cleaning any abnormal conditions, a data driven model called Similarity Based Model is proposed to capture the ideal behavior of the equipment at variable operating modes, this will then be compared to actual data and a residual signal representing the difference between ideal and actual conditions is obtained. The residual signal is compared to certain residual thresholds and departure from these thresholds flag a notification. Identification of the different fault types are performed by means of a number of rules and a classifier.

Table 2-3 Rotating and Reciprocating Equipment Vibration Limits

						Levels in mm/s RMS			Levels in micrometers		
Group #		Machine Description	Power	Speed	ISO Ref.	Acceptable	Warning	Alarm	Acceptable	Warning	Alarm
1	E.M	Small Electric Motors/ Generators	<15kW	Any	10816-1	0.71	1.8	4.5	NA	NA	NA
2	E.M	Medium size Electric Motors/ Generators	15-300kW	Any	10816-3	2.3	4.5	7.1	37 RMS	71 RMS	113 RMS
3	E.M	Large size Electric Motors / Generators	>300kW	Any	10816-3	3.5	7.1	11	45 RMS	90 RMS	140 RMS
4	Steam Turbines	Small Steam Turbines	<300kW	Any	10816-3 7919-3	2.3	4.5	7.1	37 RMS	71 RMS	113 RMS
5	Steam Turbines	Medium size Steam Turbines	300kW-50MW	Any	10816-3 7919-3	3.5	7.1	11	45 RMS	90 RMS	140 RMS
6	Steam Turbines	Large size Steam Turbines I	>50MW	1500	10816-2 7919-2	2.8	5.3	8.5	100 pk-pk	200 pk-pk	320 pk-pk
7	Steam Turbines	Large size Steam Turbines II	>50MW	1800	10816-2 7919-2	2.8	5.3	8.5	95 pk-pk	185 pk-pk	290 pk-pk
8	Steam Turbines	Large size Steam Turbines III	>50MW	3000	10816-2 7919-2	3.8	7.5	11.8	90 pk-pk	165 pk-pk	240 pk-pk
9	Steam Turbines	Large size Steam Turbines IV	>50MW	3600	10816-2 7919-2	3.8	7.5	11.8	80 pk-pk	150 pk-pk	220 pk-pk
10	Gas Turbines	Small Industrial Gas Turbines	< 3MW	Any	10816-3 7919-3	3.5	7.1	11	45 RMS	90 RMS	140 RMS
11	Gas Turbines	Medium to Large size Gas Turbines I	> 3MW	3000	10816-4 7919-4	4.5	9.3	14.7	90 pk-pk	165 pk-pk	240 pk-pk
12	Gas Turbines	Medium to Large size Gas Turbines II	> 3MW	3600	10816-4 7919-4	4.5	9.3	14.7	80 pk-pk	150 pk-pk	220 pk-pk
13	Pumps	Small size Pumps with REB	<15kW	Any	10816-1	0.71	1.8	4.5	NA	NA	NA
14	Pumps	Medium size Pumps with REB & JB	15-300kW	Any	10816-3	2.3	4.5	7.1	37 RMS	71 RMS	113 RMS
15	Pumps	Large size Pumps with REB	>300kW	Any	10816-3	3.5	7.1	11	45 RMS	90 RMS	140 RMS
16	Pumps	Large size Pumps with JB	>300kW	1000-30000	10816-3 7919-3	3.5	7.1	11	45 RMS 150-28 pk-pk	90 RMS 280-52 pk-pk	140 RMS 410-77 pk-pk
17	Compressors	Medium size Centrifugal Compressors	15-300kW	Any	10816-3	2.3	4.5	7.1	37 RMS	71 RMS	113 RMS
18	Compressors	Large size Centrifugal Compressors	>300kW	Any 1000-30000	10816-3 7919-3	3.5	7.1	11	45 RMS 150-28 pk-pk	90 RMS 280-52 pk-pk	140 RMS 410-77 pk-pk
19	Compressors	Medium size Rotary Screw Compressors	15-300kW	Any	10816-3	2.3	4.5	7.1	37 RMS	71 RMS	113 RMS
20	Compressors	Large size Rotary Screw Compressors	>300kW	Any	10816-3	3.5	7.1	11	45 RMS	90 RMS	140 RMS
21	Compressors	Reciprocating Compressors & IC	>100kW	Any	10816-6	7.07	17.8	28.2	113 RMS	283 RMS	448 RMS
22	Fans	Small Fans & Blowers	<15kW	Any	10816-1	0.71	1.8	4.5	NA	NA	NA
23	Fans	Medium size Fans & Blowers	15-300kW	Any	10816-3	2.3	4.5	7.1	37 RMS	71 RMS	113 RMS
24	Fans	Large size Fans & Blowers	>300kW	Any	10816-3 7919-3	3.5	7.1	11	45 RMS 150-28 pk-pk	90 RMS 280-52 pk-pk	140 RMS 410-77 pk-pk
25	Others	Centrifuges			10816-1	4.5	7.2	10.8			
26	Others	Gearboxes	Same as Motors classification above								
27	Others	Slow Speed Machines									
28	Others	Cooling Towers				4.5	6.3	7.6			
29	Others	Chillers				2.3	4.5	7.1			

Narrow band alarms are normally applied in the frequency domain. A number of narrow frequency bands are configured depending on specific faults of interests that are likely to occur. In the low frequency range, rotor related faults like unbalance, misalignment and looseness are likely to show up, and whereas for bearings and gearbox related problems those will likely show up at an early stage in the higher frequency ranges (Berry 1992) and (Entek).

Western Electric Company was one of the first leaders in statistical alarms (Western Electric Company 1956). These alarms are calculated based on certain statistical parameters (Mean or average, Variance, Standard Deviation). A number of measurement points are needed to enable calculating these alarms. Once an abnormal behavior is detected the next step is to identify and isolate the source of this fault. This will be covered in the next section 2.2.3.2.

Figure 2 Fault detection and identification process

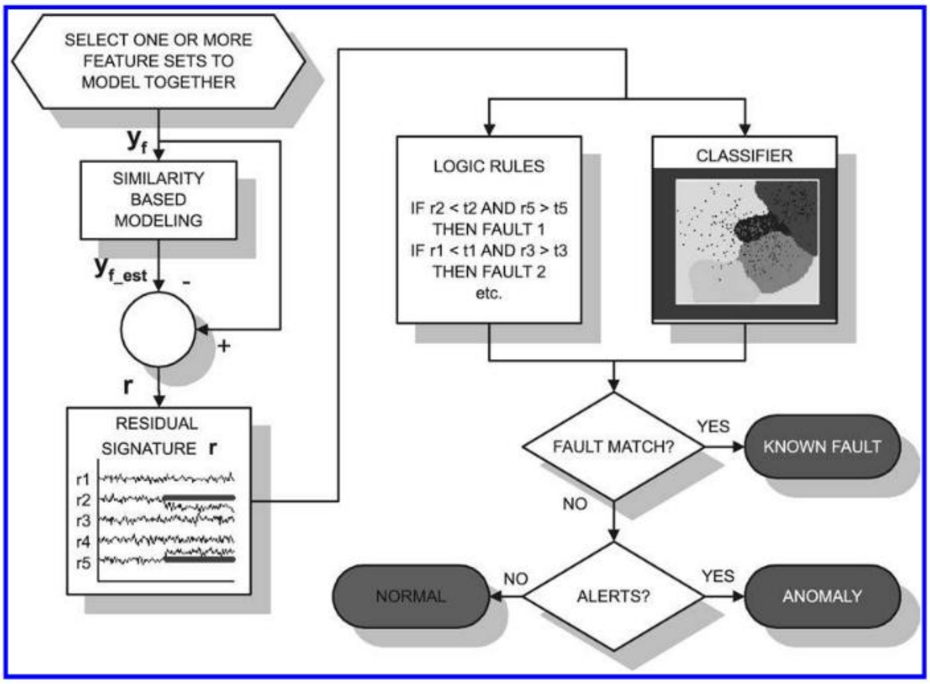


Figure 2-4 Fault Detection and Identification Process (Wegerich 2005)

2.2.3.2 Fault Isolation

Fault diagnosis is traditionally and still currently manually performed by subject matter experts through interpreting different data formats like time waveforms, trends, spectrums, bode plots, etc. for indications of fault conditions. Moving towards automation is a necessity as the whole process is limited by the availability of a competent resource to interpret the data, the time required to take actions versus doing the analysis manually and the sophisticated nature of the fault signature itself. Fault diagnostics is a rich area in the literature, hundreds of papers had been published to explore different technologies and techniques to enable better diagnostics accuracy at an early stage of fault development.

As described in (Jardine et al. 2006). Diagnostics and classification approaches are classified into three main approaches:

- **Statistical Approaches**: Hypothesis Testing Theory, Clustering Analysis using different algorithms and distance functions, Supper Vector Machines

(SVM), Convex Hull Classifier (CHC), Relevance Vector Machine (RVM), Linear Discriminant Classifier (LDC), Quadratic Discriminant Class (QDC), Hidden Markov Model (HMM), Decision Tree, etc.

- **Artificial Intelligent and Soft Computing Approaches (AI and SC)**: Multilayer Perceptron (MLP), Back Propagation Neural Network (BPNN),Self-Organising Maps (SOM), Radial Basis Functions (RBF), Fuzzy C-means Clustering, Probabilistic Neural Network (PNN), Deep Belief Neural Network (DPNN), Fuzzy Logic (FL), NeuroFuzzy Models (NF) like the Adaptive NeuroFuzzy Inference System (ANFIS), Recirculation Neural Network (RNN), Genetic Algorithms (GA), Mahalanobis Taguchi System (MTS), etc.
- **Others including Hybrid Approaches and model based approaches**

The simplest form of statistical classification problems in the field of machinery diagnostics and fault isolation is to detect the existence of a fault or not. In (Ma and Li 1995) an algorithm is proposed based on the hypothesis testing theory to test the existence of a fault in a roller element bearing by analysing the vibration data collected at variable operating conditions from good and defected bearings. The results were promising however a shortcoming for this approach was when the developed fault is small enough to have a significant observable effect, i.e. signal to noise ratio is very small.

Clustering Analysis is another statistical methods that can be applied to classification problems. This method finds groups (clusters) within the data. It sorts data based on similarity within same group (cluster) and dissimilarity outside the same group (other clusters). Examples of similarity measures used are Euclidean Distance, Mahalanobis Distance, Cosine Distance, and Pearson Correlation, etc. More details can be found in (Kasabov 2007).

In (Artes 2007) a cluster analysis study for roller element bearing was conducted to test the performance of cluster analysis in classifying data with known classes into their actual classes. Vibration Data was collected from 40 different roller element bearings with known conditions (New, with seeded faults like: Ball or roller element defect, Inner and Outer Race Defects). Cluster analysis was capable of classifying new bearings with 100% accuracy, however variable accuracy (60-80%) was noticed on the faulty bearings classes with a number of misclassified cases. The outcome of this study is that cluster analysis was able to detect the onset of a fault with 100% accuracy (i.e. a fault exists or not), however when it comes to the actual fault class the method was less accurate.

Clustering can be done on batch mode (offline) like for example: K-means clustering (Yiakopoulos et al. 2011), fuzzy clustering (Elaine et al. 2004) and (Saravanan et al. 2009), Hierarchical clustering (Ward 1963) and constrained evolving clustering method (ECMc) (Song and Kasabov 2001a) or online, like for example: evolving clustering method (ECM) (Song and Kasabov 2001a). It can also group clusters into totally isolated clusters with crisp degree of membership (0 or 1 i.e. data either belongs to or not to this cluster) and to overlapped clusters where the degree of membership is fuzzy. Several algorithms were published in clustering for data reduction and also for classification purposes (Kasabov 2007). In the offline mode the clustering analysis assumes all data are available before starting the process and this can be in one pass (iteration) or multiples iterations to enhance the classification accuracy. As most of the critical rotating equipment are instrumented with online monitoring systems, the need for an online methods that are capable of classifying faults automatically online is vital, hence the development of online clustering methods.

This research proposes a modification to the evolving clustering method for classification to enhance the classification accuracy. To the author's knowledge this is the first time for this algorithm to be used in the field of mechanical faults diagnostics and classification using condition monitoring data.

As the main area of this research within classification is built on evolving clustering method for classification, the remaining classification models will be listed below without going into much details, example reference will be given to publications using these models too.

In (Gryllias and Antoniadis 2012) simulation data for various roller element bearing conditions were generated using a physical model describing the behaviour of the roller element bearing under a fault condition. This data was used as training data eliminating the need for actual data under various faulty conditions. The SVM model was used first to classify data into normal and faulty data, another SVM model was then used to classify the faulty data into the different faults classes. The author highlighted the following reasons for deciding to use the SVM model:

1- Supervised Learning Methods like Artificial Neural Network have the following limitations: require large dataset size for training the models, can converge to a local minima solution, has poor generalisation, can overfit the model by learning the noise also as large dataset is required.

2- Unsupervised Learning Methods like some of the clustering algorithms can't identify the exact fault class label as no labels were used during the training, only inputs. In which case a subject matter expert is still required to identify the labels.

Supervised learning methods have limitations in terms of the lack of training inputs/output datasets, changes of the number of classes, etc. However, for the cases where historical information about system health, e.g. fault modes observed, fault levels measured, are available, the supervised-learning methods are able to provide better results than unsupervised-learning methods due to the knowledge learned from the historical data.

(Li et al. 2012), (Sugumaran and Ramachandran 2011), (Younus and Yang 2012), (Jegadeeshwaran and Sugumaran 2014) and (Yuan and Chu 2006) are example publications where SVM was also used for fault classification of various components/equipment.

A comparison between CHC, LDC, QDC and MLP was made in (Volpi et al. 2010). The idea is to create a classification problem with only one class (the new condition class) as this type of data is always available. The CHC model is trained using the new condition class in order to classify any incoming data during test to either this trained class or to an unknown class. A side advantage of using one class is the small memory size required and the reduced processing time. The testing results showed that the CHC model was capable of identifying the unknown classes better than the other three models.

RVM is used in (Tran et al. 2013) for fault classification of a rotor kits using thermal imaging data. The proposed method is presented in Figure 2-5.

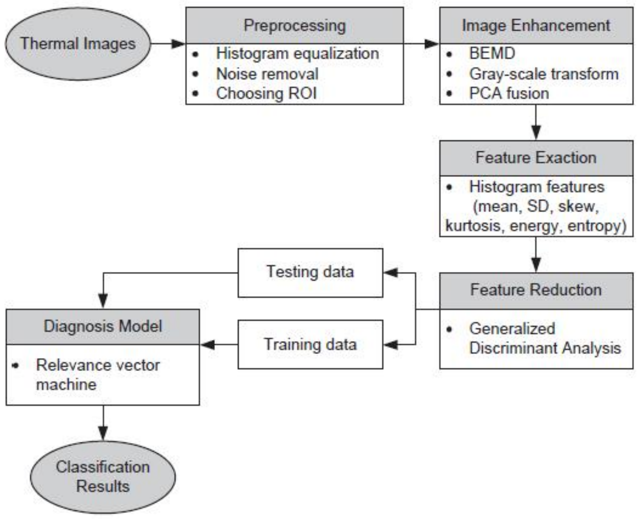


Fig. 1. Proposed system for thermal image based fault diagnosis.

Figure 2-5 Fault Classification Process Using BEMD and RVM (Tran et al. 2013)

After pre-processing the images for noise removal, etc. The bi-dimensional empirical model decomposition (BEMD) was used to decompose the image into its intrinsic modes and a residual. Statistical features were then extracted and a data reduction method by means of the generalised discriminate analysis method is performed. RVM was finally used for fault classification. A comparison between this method and SVM and ANFIS was made. All models showed similar training accuracy identifying a 100% of all cases however RVM resulted in better classification accuracy (by 2.5% and 15% respectively) compared to one of the SVM models used and ANFIS. The second SVM model showed similar testing accuracy to RVM however processing time with RVM was shorter.

HMM based model called HMMSNP (where SNP refers to semi-nonparametric) was used in (Geramifard et al. 2013) for classification of bearing and rotor faults in electric motors. The usual way of training the HMM is that for each fault there will be a trained HMM. For each new incoming data, the log-likelihood will be generated by each HMM trained to identify any similarity between this data and the data trained on each separate HMM. The proposed model was expected to perform better than HMM and the results actually showed 28% improvement over the HMM.

(Yan et al. 2011) and (Li et al. 2006) are example publications where HMM with some modifications was also used for fault classification of various components/equipment.

Acoustic data was acquired in the vicinity of a rotating kit with good and faulty bearings in (Karabadji et al. 2014). This was then followed by feature extraction and data reduction. The C4.5 decision tree was used for classification of the bearing faults. Figure 2-6 shows the decision tree used. The parameters used here are the skewness, mean and median, kurtosis, Inner Race and Outer race frequencies. The accuracy obtained was 95.5%.

(Amarnath et al. 2013), (Muralidharan and Sugumaran 2013), (Sakthivel et al. 2010a) and (Sun et al. 2007) are example publications where Decision tree with some modifications was also used for fault classification of various components/equipment.

MLP Neural Network was compared to Kohonen SOM in (Ghate and Dudul 2010). Motor current data were analysed for 4 different electrical faults. The MLP Neural Network outperformed the prediction accuracy of the SOM by ~2%, at 98.25%.

(Barakat et al. 2011) , (Wu et al. 2010), (Wu and Liu 2009), (Rafiee et al. 2007), (Wang et al. 2011), (Tran et al. 2014), (Xu et al. 2009), (Ye 2009), (Wang and Hu 2006), (Sathiyasekar et al. 2011), (Wang and Kanneg 2009), (Zhang et al. 2005), (Pan et al. 2010), (Jin and Chow 2013), (Li et al. 2013), (Deng and Zhao 2013), (Chen and Chen 2011), (Sakthivel et al. 2010b) and (Li et al. 2011) are example publications where other artificial intelligent and soft computing approaches (including NN, FL, NF, GA, fuzzy c means clustering and MTS) were used for fault classification of various components/equipment.

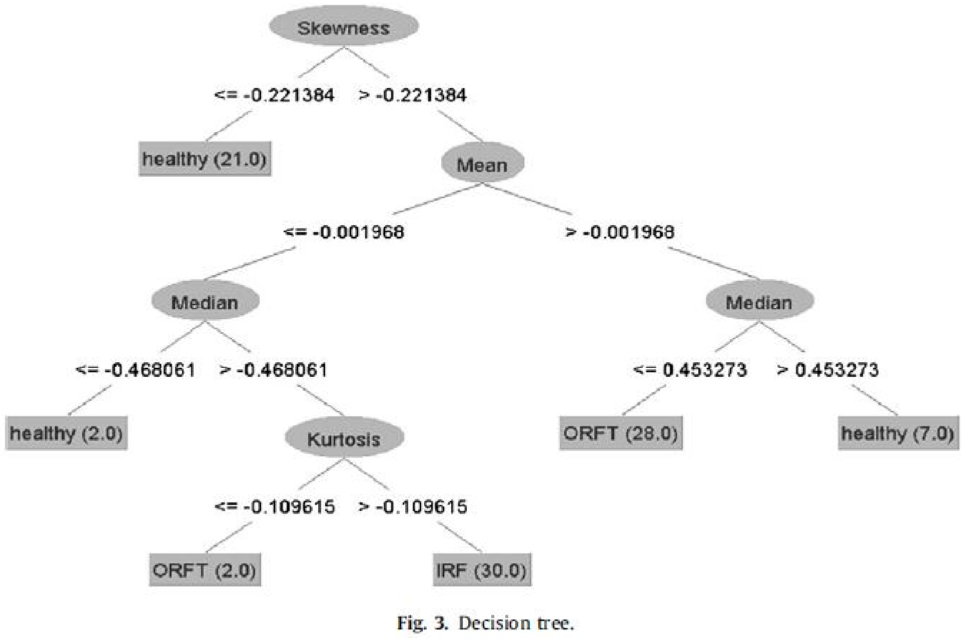


Figure 2-6 C4.5 Decision Tree (Karabadji et al. 2014)

2.2.4 Fault Progression Prediction (Prognostics)

As mentioned in section 2.1.2 there are three different approaches to prognostics:

- **Model Based (Physical Based) Approaches**
- **Data Driven Approaches**
- **Hybrid Approaches**

The above models will be described in some details below however the current research is a hybrid approach between a number of data driven approaches as such more emphasis will be given to the second and third approaches.

2.2.4.1 Physical Based Approaches

The physics based prognostics approach is a prognostics model built based on the system or component physical behaviour governed by physical laws like Fracture Mechanics or based on system identification methods. Alot of work has been published in the literature in this area with a concentration on bearings and gearbox applications compared to other applications due to the wide use of these components and mainly in the Space field (Aircrafts & space shuttles) due to the criticality and importance of these applications and the long term missions which require highly reliable systems to withstand all environmental conditions. These models aren't only concerned with the initiation of Faults but with their propagation also. Different papers have been reviewed here showing the variety of applications of the Physics based prognostics, its merits and limitations.

Paris's Equation is one of the most popular used formulas for describing the fatigue crack propagation (growth) rate, equation 2-4:

$$\frac{da}{dN} = C_o (\Delta K)^n \qquad \textbf{(2-4)}$$

Where :
a = Instantenous length of crack
N = No. of Cycles
C_o & n are constant material properties
and
ΔK = Stress Intensity Range

The above formula describes the deterministic part of the Crack propagation which is between the crack initiation time and the point at which the crack size becomes critical. This is normally called the stable zone however, beyond this point the behaviour becomes stochastic (unstable) and the propagation growth rate can no more be described in a deterministic way.

In (Li et al. 1999) the authors used Paris's equation with slight changes to reflect the crack area size growth rate instead of the crack length size to study the defect propagation rate of a rolling element bearing. Vibration data has an indirect relation to the crack size and because the exact crack size can't be measured without stopping the machine and physically examining the bearings, the vibration measurements were used to fine tune the model and to reduce the uncertainty inherent in the model as a result of the stochastic behaviour of the crack growth, having in mind that this model was deterministic.

The author(s) in (Li et al. 2000) built on the same model in (Li et al. 1999) but added a non-linear stochastic variable to the equation- a lognormal random variable such that the stochastic nature of the crack growth can be studied and modelled and a more accurate prediction can be obtained. The model was validated experimentally on rolling element bearings, results showed that the model is capable of predicting the time to reach the critical size of the crack with low uncertainty of less than ±0.4 million cycles.

A novelty stochastic model based on Fracture mechanics and Forman equation was integrated with CM & inspection data in (Kacprzynski et al. 2002) and experimentally validated using spur and helical gears in a trend to apply it to helicopter's gearboxes. The model is a 2D Finite element fracture mechanics model and acoustic emission data was collected in a run to failure test. A confidence of 97% in the predicted remaining interval/cycles was claimed.

Yu- and Harris equation (Ioannides and Harris 1985) was used in (Orsagh et al. 2003) to model the Spall initiation process. Kotzalas and Harris spall progression model was also utilised. The stochastic model was fused with oil debris and vibration data collected on gas turbine engine bearings, the objective was to develop a prognostics model for the oil wetted gas turbine engine components. Bayesian and Demspter Shafer methods were used to minimise the uncertainties in the model predictions. The vibration features and oil debris were fused on the features level. the Yu-Harris bearing life equation in its stochastic form is used in (Orsagh et al. 2004) for bearing fault initiation prediction in a gas turbine application. The model was integrated with vibration data.

The singular perturbation method coupled with dynamic state estimation was utilised in (Luo et al. 2003a) to build a prognostics model under variable loading condition the application was an automotive suspension system. The authors also reviewed the available prognostics techniques.

Continuum Mechanics and FEM were used to model the fatigue damage in a gas turbine blade (Kumar et al. 2007). The idea was to compare the changes of frequencies as a result of the blade stiffness changes due to the defect growth with certain thresholds that are indicators of damage.

A mathematical model for half car suspension system was built in (Luo et al. 2008). The singular perturbation method coupled with dynamic state estimation techniques were used for the prognostics model of the suspension system.

An interesting application for physics prognostics models in monitoring the diamond turning tool chipping was studied in (Choi and Choi 1996) in order to maintain a high quality and precision for the surface finish. Measurement of the force signal was utilised in the model to describe the surface profile and the model was based on the minimum cross entropy and Kullback's principle. Experimental validation for the model was conducted using fresh and chipped tools for comparison purposes.

In (Kumar et al. 2007) the author(s) studied an integrated physics based prognostics model with CM data for aero engine components. The paper highlighted some of the deficiencies and problems with using physics prognostics models independently:

1- Need intensive computations that might be expensive also.
2- The stochastic nature of health degradation rate causes large uncertainties in these models prediction. In addition, these models are normally validated experimentally in ideal conditions compared to reality.
3- Limitations and effectiveness to new structural systems.
4- Unreliable in real time/actual life assessment of components.

Other physics prognostics models using fracture mechanics and system identification methods can be found in (Luo et al. 2003b), (Jia et al. 2003), (Sekhar 2004), (Abdel-Magied et al. 1998) and (Li and Lee 2005).

Physics-based prognostics approaches have many limitations in terms of:

1- The complexity of the models i.e. not practical
2- The fact that these models are component and fault specific i.e. not generic.
3- Obtaining high predicted accuracy can be achieved with these models but with expensive methods like FEM.
4- A number of assumptions are made at the beginning that makes them describing ideal rather than real cases.

2.2.4.2 Data Driven Approaches

Data driven prognostics models are depending on the CM collected data and failure (historical) data to train the prognostics model and to predict the remaining useful life. This approach can be described generally as a model free approach where no mathematical physical model is needed which makes this approach more practical and easier to be applied reducing the amount of uncertainty caused by the simplified assumptions proposed at the beginning of building the physics models. However this approach has a number of limitations:

1- It requires high quality and large quantity of data.
2- Some of these models are unable to predict the transient changes in the component life line.

Three general categories of methods are mentioned in the literature for the Data-Driven approaches:

1- Projection (trend analysis) Statistical models: Reliability Models, Exponential smoothing, AR, ARMA and ARIMA, etc.

2- AI and Soft Computing Models: NN, Fuzzy Logic, NF, GA, Evolving Connectionist methods, etc.

3- Stochastic & Probability Models: Hidden & Semi-Hidden Markov Models, Bayesian approach.

The above methods are very well established and developed in diagnostic applications in addition to the pattern recognition techniques and clustering methods.

In prognostics, there are two types of predictions: predicting the remaining useful life and predicting whether the equipment will survive to finish a specific task or until the next planned shutdown. Few publications were concerned with the second type of prediction. The two prediction types are referred to in this research as short term and long term predictions.

A reliable machine is a machine that can operate properly for a certain period of time without failure. Reliability Analysis is concerned with the population of equipment that have common characteristics in the plant rather than with individual equipment. The most important piece of information in this regard is the time-to-failure that can be obtained from the maintenance history of these machines, the time to failure for an individual component can be calculated accordingly (Brown and Meyer 1961).

Trending and projection methods are based on fitting a series of data to a specific curve (ex. Line or polynomial) and extrapolating (estimating) future readings. Exponential Smoothing was first introduced in (Mentzer 1988) for data presentation and time series prediction methods that are used for smoothing the data and assigning exponentially decreasing weights to the previous readings in order to forecast future data. One of the popular applications for exponential smoothing is in marketing and sales and inventory control management applications (Batko 1984). (Usynin 2006) is an example of another application for the exponential smoothing in the diagnostics field. A special form of the exponential smoothing where the previous historical data are equally weighted is the moving average. These methods are very simple and easy to implement however there is a high uncertainty in the prediction especially at times of sudden changes, also the effect of historical outliers readings are very bad on the predicted future readings.

In (Usynin 2006) degradation data were fitted to different models including the cumulative wear, logistic and nonlinear regression models to provide an estimate for the future condition of the components, however an assumption was made at the beginning that the degradation is a deterministic one which isn't quite right specially that some of the degradations are stochastic in nature and future condition isn't related to past behaviour of the machine.

Artificial Neural Networks were originally developed to describe artificially how the mind is working. In practice, they are nonlinear networks, an assembly of layers interconnected with multi-input, hidden and output layers. The inputs are normally CM data, Failure rates, etc. And the output is the predicted remaining useful life. Figure 2-7 shows a simple Neural Network, sometimes called Feed Forward (Static) Neural Network (FFNN). It's called Feed Forward Neural Network because the inputs to the layers are going in one direction forward (No feedback).

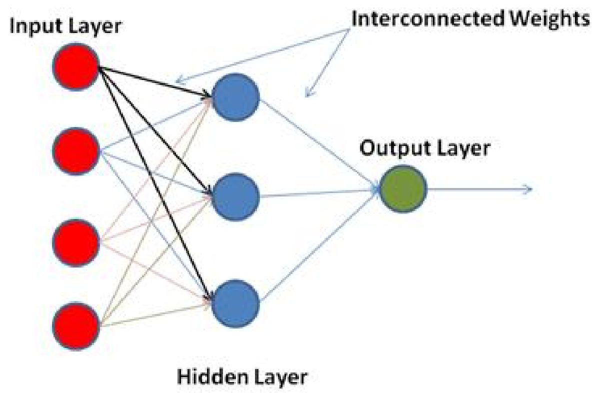


Figure 2-7 Feed Forward Neural Network

Recurrent Neural Network (RNN) was used in (Tse and Antherton 1999) to forecast the machine deterioration rate using vibration data. The RNN is a closed loop NN with the interconnected weights going in both directions. This is to accommodate for the dynamic behaviour of the system/Component degradation. Results of testing from (Tse and Antherton 1999) showed that the RNN outperforms the other time series trending methods including the FFNN and the Autoregressive Moving Average ARMA methods.

A comparison between the RNN and NF (Neuro-Fuzzy) system was studied in (Wang et al. 2004). Data from a chipped, worn and cracked gears are used to train both RNN and NF networks. The Neuro-fuzzy system is a system integrating NN and Fuzzy Logic in a single model since the fuzzy logic model doesn't have the capability to learn, the NN will do the learning part and the reasoning is performed by Fuzzy Logic. The output from the study showed that the NF performs better than RNN when sufficient data are used for the training with higher accuracy.

Regression tree is a nonlinear prediction model that uses CM and failure data to construct a decision tree which is used then to classify new data. It's a classification method like data clustering. Questions that have a yes/no answer are normally asked to split the learning data into smaller and smaller parts (Timofeev 2004). In (Cao 1997) the regression tree method (CART Classification And Regression Tree) was used as a prognostics model to perform one step ahead prediction for a methane compressor using real vibration data collected over three months up to failure in addition to historical data. The model can perform multistep ahead predictions however the more steps ahead the higher uncertainty in the predicted values. The Cao's method (Cao 1997) was used to calculate the embedding dimension which is the number of needed inputs for the above model.

One problem of CART is that it doesn't have the capability to predict the sudden changes in the system as found in (Tran et al. 2009) where two models: CART and Adaptive Neuro-fuzzy system were studied for comparison purposes using the same compressor vibration data from (Cao 1997). Three strategies for obtaining a multi-step ahead prediction were discussed, Recursive, DirRec and Direct prediction strategy. The last one was used here for multi-step ahead prediction as it has the highest prediction accuracy because it avoids the accumulation of errors from previous prediction runs.

A library of time to failure cases with different fault patterns were developed and the fuzzy similarity analysis was utilised for the remaining useful life prediction by comparing new data to the data stored in the library through their similarity in developing the failure patterns (Zio and Maio 2010).

In (Vachtsevanos and Wang 2001) the authors proposed a generic prognostics approach using Dynamic (Recurrent) Wavelet Neural Networks DWNN. The proposed model was validated using vibration data. The DWNN is a NN with classifications capability obtained by integrating wavelets to the NN. The model is capable of utilizing vibration signals, acoustic data, process parameters like temperature and pressure, historical failure data. The DWNN can be coupled to one of the probability methods like Dempster-Shafer to manage the uncertainties coming from the prognostic method, human errors and from the process itself.

Monte Carlo method (Particle Filter) was used with real vibration data from a methane compressor for remaining useful life estimation and for crack growth estimation (Caesarendra et al. 2010) and (Cadini et al. 2009). Extended Kalman Filter was used in (Shao and Mechefske 2009). Note that Monte Carlo methods are optimal for nonlinear models whereas Kalman Filter is used for linear models.

Traditional statistical forecasting methods are very simple to be applied when dealing with simple cases; when the independent and dependant variables are well defined however they aren't able to deal with nonlinear complex relationships between variables and sudden dynamic changes. Soft computing and AI techniques on the other hand are able to infer complex nonlinear relationships purely based on the data without any prior assumptions in the model. Each one of these methods has its own shortcomings, however a hybrid between different AI techniques has shown better performance.

For example, a hybrid method between NN and Fuzzy Logic like ANFIS (Adaptive NeuroFuzzy Inference System) (Jang 1993) which is considered one of the most popular NeuroFuzzy Inference methods to date (Kasabov 2007) gave a better option combining the learning and adaptation capabilities of Neural Networks with the IF-THEN rules reasoning of fuzzy logic to find the complex inputs/output nonlinear relationships with a better visibility into this network black box reasoning. ANFIS has some limitations in terms of the cases where it can be applied though as it isn't able to handle cases where the number of inputs is high due to the curse of dimensionality. The structure of the ANFIS network is fixed one meaning that the network will not be capable of adapting to any new data and has limited online capability (Kasabov 2007).

Evolving connectionist methods are very promising with their ability to adapt to new data, incremental learning, open structure to add new features and rules and most important can start learning with minimum amount of training data. DENFIS (Kasabov and Song 2002) is one of the evolving connectionist methods with the grid partitioning of the input space done using a novel evolving clustering method (ECM), and a recursive least square method distance weighted for solving the unknowns (consequent parameters) of the output equations. Fuzzy Inference System (FIS) is the processing engine in DENFIS and the type of fuzzy rules is a 1st order Takagi-Sugeno ones (Takagi and Sugeno 1985). The method was compared with other online learning methods like RAN (Platt 1991), EFuNN (Kasabov and Song 2002), ESOM (Kasabov and Song 2002) and better testing NDEI and processing times were found.

The DENFIS model can be used also in an offline mode using the constrained evolving clustering method following the use of the online evolving clustering method. This research will study different modifications to the original model and their effect on the prediction accuracy of benchmark datasets and field data. In addition to utilizing a global optimization method – Particle swarm optimization (Takagi and Sugeno 1985) with effective local approximation method (Lin et al. 2006) to optimize the solution obtained by the weighted recursive least square method.

3 DESIGN OF A NOVEL HYBRID CONDITION MONITORING AND PROGNOSTIC MODEL

3.1 Introduction

This chapter presents a novel hybrid condition monitoring and prognostic model for addressing some of the identified issues from the literature that aren't yet resolved. The proposed model takes advantage of a hybrid between Neural Network and Fuzzy Logic model which has the capability of learning the nonlinear relationships between a number of condition monitoring parameters and an output which represents the health indicator of interest. The proposed model also utilises one of the global optimisation solutions to enhance the accuracy of the model's predictions.

With minor changes in the model, it can be used for a number of applications which as a package offer a complete integrated solution for condition monitoring and prognostics of rotating equipment. The proposed model is able to predict the healthy behaviour of the condition monitoring parameters taking into consideration variations in the loading and environment conditions. It also notifies the end user of any departure from the normal operating envelope at which point an automated diagnostic system, also proposed as part of the integrated solution, will process the data to identify and isolate the root of the problem (diagnostics). The proposed model is capable of producing short term and long term predictions.

To better understand the proposed model, section 3.2 presents a full review of all the enabling techniques theory. Section 3.3 explains in details the architecture of the proposed condition monitoring and prognostic framework and a summary for chapter 3 is provided in section 3.4.

3.2 Enabling Techniques for the Development of the Proposed Model

3.2.1 Takagi-Sugeno-Kang (TSK) NeuroFuzzy Network

A neuro-fuzzy network is a hybrid between two different techniques namely neural network and fuzzy logic. It takes advantage of the adaptive learning and reasoning capabilities of neural networks and fuzzy logic, respectively. The parameters of the fuzzy system in terms of membership functions (MFs) and fuzzy rules are determined using neural network learning algorithm with input/output pairs. The first studies in this area are going back to the 90s, (Jang 1992), (Jang et al. 1997), (Jang 1993) and (Lin and Lee 1991). A brief overview of neural networks and fuzzy logic is presented here to give some background about the origins of neuro-fuzzy network.

3.2.2 Neural Network

Neural network is defined as a set of interconnected nodes that simulates the human brain reasoning in order to understand complicated systems and solve their problems without the need to develop an actual model of the system, having in mind that complicated systems are hard to be modelled using classical mathematical equations. They are purely data driven models. Initial work in this area was done by (Bain 1873) and (James 1890) but due to the need for more advanced technologies to validate any new ideas by simulation not much work was done until the development of the first single-layer perceptron, the simplest type of feed forward neural networks (Rosenblatt 1957), however it was only single layer perceptron which has the capability of learning the linear relationships between inputs and output, Figure 3-1. Recent developed neural network are multi-layer Networks.

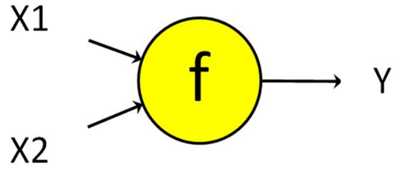


Figure 3-1 Single Layer Feed Forward Network

In (Jang et al. 1997) neural networks are classified based on the following, see Figure 3-2:

1. Learning Method
2. Architecture
3. Output Type
4. Node Type

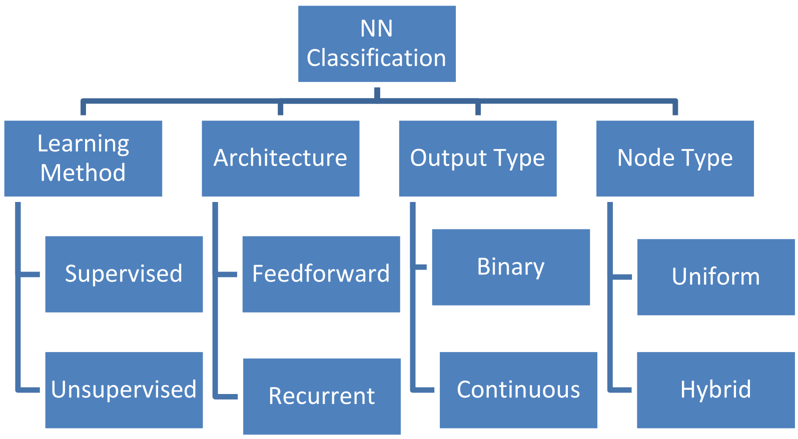


Figure 3-2 Neural Networks Classifications

Adaptive networks are networks that consist of a number of nodes and links between them. The nodes can be adaptive with modifiable or fixed function. The links between the nodes explain the relation between them. These links can have a weight or can be weightless. Figure 3-3 shows the different shapes of nodes found in a typical network. The square shape is for adaptive nodes while the circle shape is for fixed nodes that contain no parameters set. The arrow is for the directional relation between interconnected nodes, these links can be forward and in one direction or both forward and backward in a recurrent network.

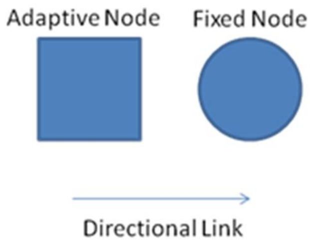


Figure 3-3 Adaptive Network Legend

The below network, Figure 3-4, is a single layer feed forward network which contains one adaptive node f, the output of this network Y can be described by the two inputs X1, X2 and the node parameters as follow, equation 3-1:

Figure 3-4 Single Layer Feed Forward Network Example

$$Y = f(X_1, X_2; a_1, a_2, a_3) = a_1 X_1 + a_2 X_2 + a_3 \quad \textbf{(3-1)}$$

The only limitation for the type of function used in the network is that it must satisfy the piecewise differentiability requirement which means that each piece of this function must be differentiable throughout its entire domain.

3.2.3 Fuzzy Logic

Fuzzy logic was presented in 1965 (Zadeh 1965) to explain the vagueness in some applications. In classical sets the boundaries are clear and crisps (precise) like 0 or 1, true or false; whereas in Fuzzy sets the boundaries are gradual or smooth and imprecise that vary between 0 and 1. For example with the weather forecast the weather can vary from being extremely hot to extremely cold with different regions in-between rather than just cold or hot, this is including expressions like fairly cold, slightly warm, etc. Fuzzy sets are best fit in these situations. The inputs are mapped by all functions that cover different ranges of the domain to find out the true values of the outputs. For example:

- “IF Premise (Antecedent), THEN conclusion (Consequent)”. (Ross 2004)
- IF Vibration IS very low THEN invalid reading
- IF Vibration IS low THEN reading acceptable
- IF Vibration IS moderate THEN reading is still acceptable

- IF Vibration IS high THEN investigate the problem
- IF Vibration IS very high THEN stop the machine and investigate the problem

The MFs can have a value between 0 and 1. These functions map all the inputs to each membership grade. A common type of MFs is the ***generalized bell-shaped*** MF. An example of the generalized bell-shaped MF is shown in Figure 3-5. This MF depends on the following parameters: a (width) , b (slope) at $f(x)=0.5$ and c (centre) as per the below equation 3-2.

$$f(x;a,b,c)=\frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \quad \textbf{(3-2)}$$

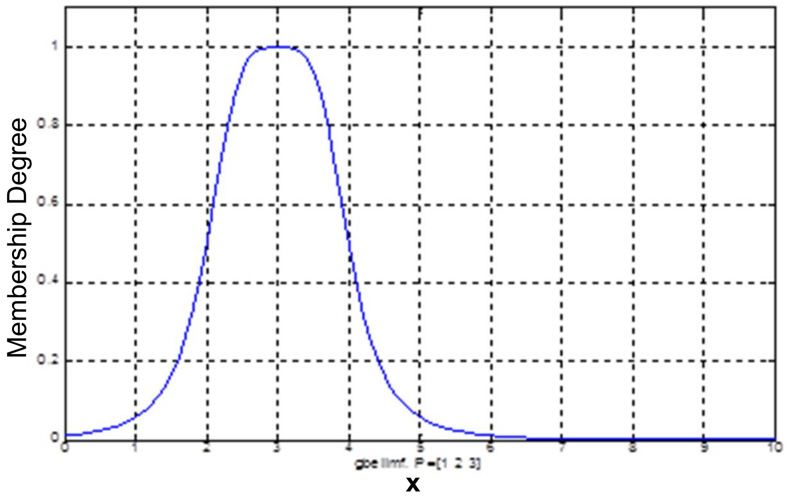


Figure 3-5 A Generalised Bell Shaped MF

Varying the values of the width at constant slope and centre and the slope at constant width and centre are shown in Figures (3-6) & (3-7) below.

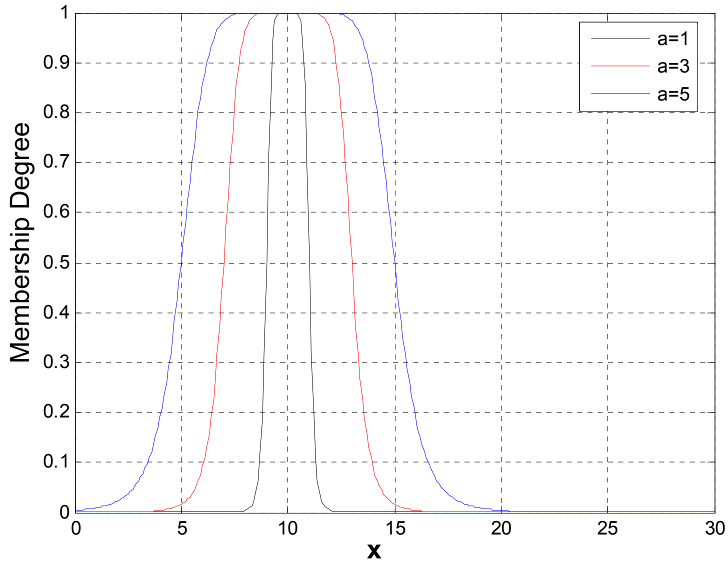


Figure 3-6 A Generalised Bell-Shaped MF at Various Width Values

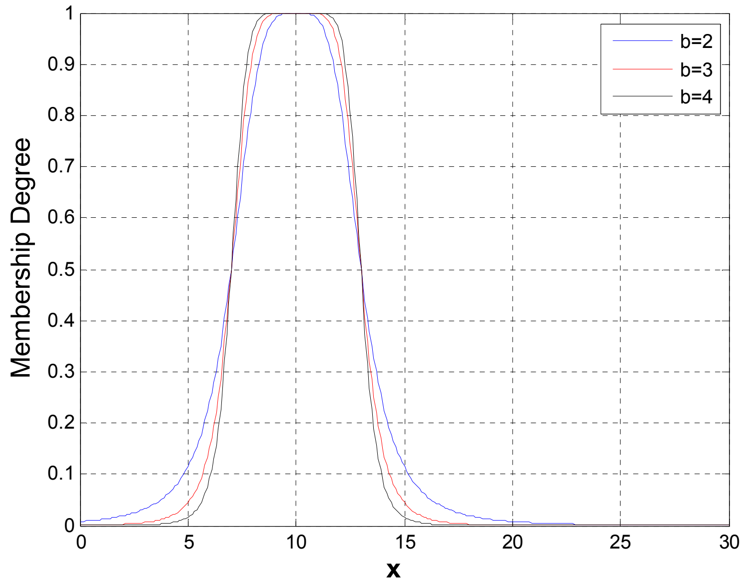


Figure 3-7 A Generalised Bell-Shaped MF at Various Slope Values

3.2.4 Fuzzy Inference Systems

A mapping between inputs and outputs is created using Fuzzy Inference Systems (FIS) using fuzzy logic rules. FIS has various applications in different fields including: time series forecasting, expert systems, pattern recognition...etc. (Jang et al. 1997).

FIS have two main types that are used in the literature:

1. Mamdani FIS (Mamdani and Assilian 1975).
2. Takagi-Sugeno-Kart (TSK) FIS (Sugeno 1985).

Mamdani used human expertise to develop a number of fuzzy rules in order to control a steam engine and boiler. Two rules were used to develop an output; Defuzzification was implemented to transform the output from fuzzy to crisp which then can be used by the facility. Examples of Mamdani fuzzy rules are:

- IF Vibration IS low THEN reading acceptable
- IF Temperature IS high and Vibration IS high THEN bearing damage is imminent

In the TSK FIS method the rules are somehow different. A more systematic approach to the fuzzy rules development was proposed. Example of a TSK rule:

- IF x IS X1 and y IS Y1 THEN $z = f(x, y)$, X1 and Y1 are two fuzzy sets and z is a crisp function.

z is a crisp function of the inputs in a polynomial or other form that describes the output. Example is shown in equation 3-3.

$$z = f(x, y) = v_1 x + p_1 y + a \tag{3-3}$$

When the function is in the above format i.e. equation 3-3, the FIS is called first order TSK FIS. A special case is when f =constant in which case the FIS is called zero order TSK FIS (Mamdani FIS). The main difference between both FIS models is to do with the consequent constituent. The output from the TSK FIS is either polynomial or constant (crisp output) while the output from Mamdani FIS is fuzzy output (Wang 1994).

In the case of Mamdani FIS a defuzzifier is required to get a crisp output. The

Mamdani FIS has an interpretable consequent part of the rules (linguistic) while the consequent part of the rules with the TSK FIS isn't linguistic i.e. hard to be interpreted. However with the TSK FIS the consequent part of the rules has the number of parameters equals to the number of input variables i.e. more degrees of freedom and it is continuous, meaning that for each input there is only one output adding to this that its processing time is shorter than the Mamdani FIS since with the later Defuzzification is required (Mendel 2001).

3.2.5 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a method first proposed in (Jang 1993) integrating both fuzzy logic (Zadeh 1965) and neural network (Ross 2004) to a more robust solution for modelling and forecasting. This method utilizes the capability of the fuzzy logic reasoning in the form of IF-THEN rules to form a set of constraints from the training data to reduce the optimization search space hence finding a model that describes the complex nonlinear relationships among a system that conventional mathematical differential equations can't describe.

Neural Networks have the learning capability especially that it can adjust itself to any type of behaviour or situation (Adaptive) based on experience during the learning stage. Fuzzy logic and neural network have some common features like the fact that both of them assume parallel operation and estimate functions purely from sampled data (model free).

ANFIS architecture is similar to the first order Takagi-Sugeno-Kang (TSK) FIS. Figure 3-8 shows a first order ANFIS model with 3 inputs (x, y, z) and a single output (O). Each input has two membership functions (Layer 1) total of 3x2=6 MFs. The number of IF-THEN rules generated in this case is 2^3=8 rules. One of the limitations of this model is the curse of dimensionality which occurs when the number of variables (inputs) is very large i.e. if for an instance the number of variables is 15 then the number of IF-THEN rules generated will be 2^{15}=32768 rules this is computationally very expensive and will take a very long time leading to an out of memory error. There are 5 layers in the ANFIS model, the 1st and 4th are adaptive nodes the rest are fixed nodes. The parameters of the 1st layer are called premise (antecedent) parameters while the 4th layer's parameters are called the consequent parameters.

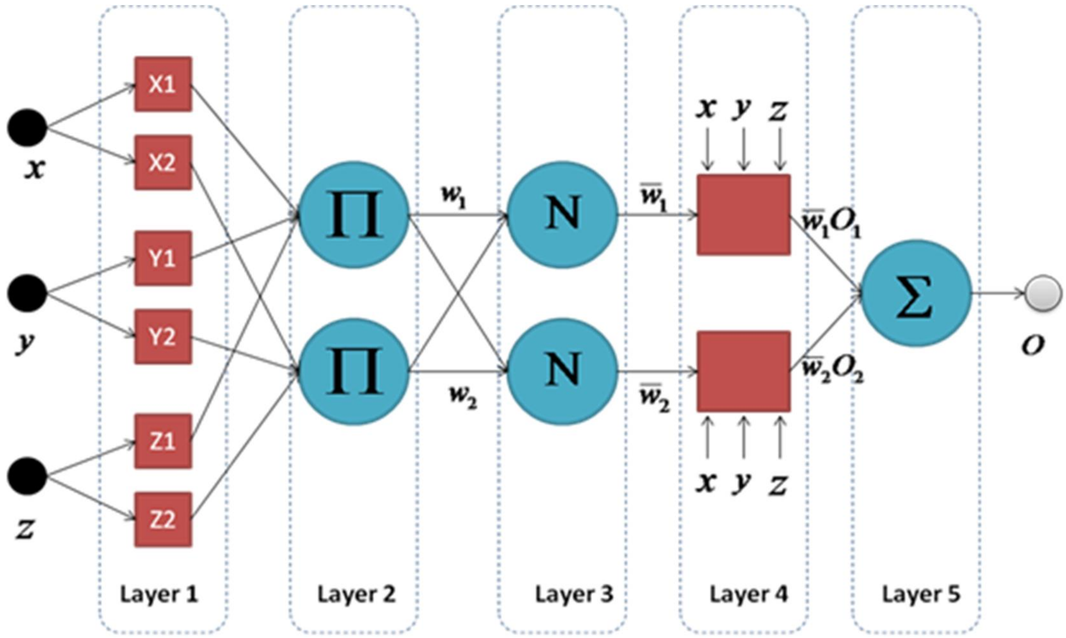


Figure 3-8 First Order TSK ANFIS Network

Layer-1: Contains adaptive nodes that have functions similar to the generalized bell-shaped membership function mentioned in section 3.2.3 used to map crisp input values to fuzzy linguistic expressions and vice versa. Equation 3-4 describes this MF.

$$\mu_X(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}} \tag{3-4}$$

$$M_{1,i} = \mu_{X_i}(x) \text{ where } i = 1,2$$
$$M_{1,i} = \mu_{Y_{i-2}}(y) \text{ where } i = 3,4$$
$$M_{1,i} = \mu_{Z_{i-4}}(z) \text{ where } i = 5,6$$

x, y, z are the inputs to this model
X, Y, Z are linguistic terms
a_i, b_i, c_i are the premise parameters used to define the fuzzy region

Other types of MFs that can be used including:

1. Gaussian curve MF
2. Gaussian combination MF
3. П-shaped built-in MF
4. Product of two sigmoidal shaped MF
5. Difference between two sigmoidal MF
6. Trapezoidal shaped MF
7. Triangular shaped MF

(Zhang and Bose 2002) Studied using different types of MFs in an induction motor fuzzy controller. The results showed that the triangular and trapezoidal shaped MFs give the best results for this type of application in terms of speed overshoot (none), speed response (fast), steady state accuracy (high) in addition to other KPIs. The polynomial PI MFs gave the worst performance.

<u>Layer- 2</u>: Contains fixed nodes i.e.no modifiable parameters. The output is the product of all input data to these nodes, described in equation 3-5.

$$w_i = \mu_{X_i}(x).\mu_{Y_i}(y).\mu_{Z_i}(z) \text{ where } i = 1,2 \quad \text{ith rule firing strength} \qquad \textbf{(3-5)}$$

<u>Layer-3</u>: Contains fixed nodes i.e.no modifiable parameters. The output of this layer is the ratio between each i^{th} rule firing strength and the sum of all firing strengths (normalization), equation 3-6.

$$\overline{w}_i = \frac{w_i}{w_1 + w_2} \text{ where } i = 1,2 \qquad \textbf{(3-6)}$$

<u>Layer-4</u>: Contains adaptive nodes. The output from this layer is the product of the i^{th} normalized firing strength and a polynomial function that has the consequent parameters, equation 3-7.

$$O_{4,i} = \overline{w}_i.O_i = \overline{w}_i.(v_i x + p_i y + q_i z + r_i) \text{ where } i = 1,2 \qquad \textbf{(3-7)}$$

v_i, p_i, q_i and r_i are the consequent parameters used to identify the output in the fuzzy region

<u>Layer-5</u>: Contains one fixed node. The output from this node is the sum of all inputs to this node being the estimated output from this model which will later be compared to the actual output in order to evaluate the prediction accuracy, equation 3-8.

$$Output = O_{5,1} = \Sigma \overline{w}_i . O_i \quad \text{where } i = 1,2 \tag{3-8}$$

The premise (nonlinear) and consequent (linear) parameters are learned in a hybrid learning process, described in Figure 3-9 using the gradient descent and least square methods. The output and error are estimated in the forward pass and backward pass respectively.

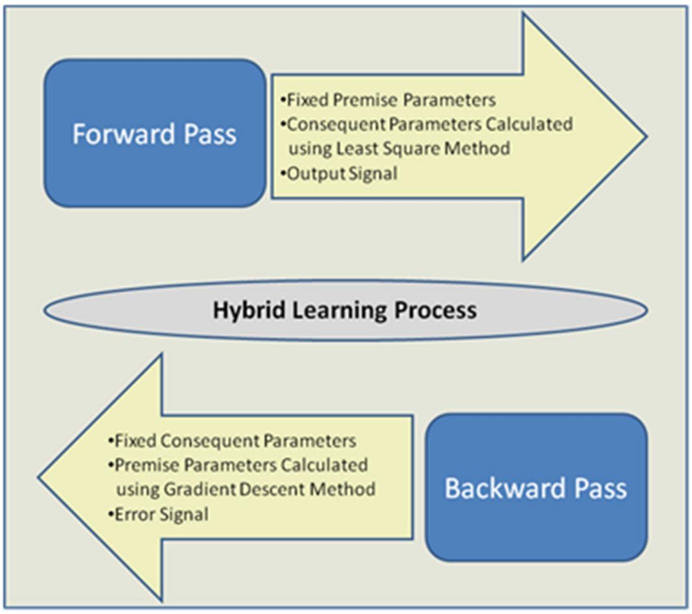


Figure 3-9 ANFIS Hybrid Learning Process

3.2.6 Weighted Recursive Least Square Method

The output of a linear model y is a linear combination of the inputs $x_1, x_2, \ldots, x_q$ in the following form, equation 3-9:

$$y = b_1 x_1 + b_2 x_2 + \ldots + b_q x_q + b_0 \qquad \textbf{(3-9)}$$

Where $b_0, b_1, b_2 \ldots b_q$ are the unknown parameters to be estimated using inputs/output pairs. The least square method is used to solve this equation by minimising the squared error between the actual and estimated output. The least square (LS) solution to equation 3-9 is shown in the below equations:

$$y = Ab \qquad \textbf{(3-10)}$$
$$\hat{y} = A\hat{b} + e$$
$$e = \hat{y} - A\hat{b}$$
$$E = \text{Sum of Square Error} = e^T e$$
$$E = \left(\hat{y} - A\hat{b}\right)^T \left(\hat{y} - A\hat{b}\right) = \left(\hat{y}^T - A^T \hat{b}^T\right)\left(\hat{y} - A\hat{b}\right)$$

$$\text{To minimise the sum of squared error, assume } \frac{\partial E}{\partial \hat{b}} = 0 \qquad \textbf{(3-11)}$$
$$E = \hat{y}^T \hat{y} - 2\hat{b}^T A^T y + \hat{b}^T A^T A\hat{b}$$
$$\frac{\partial E}{\partial \hat{b}} = 2A^T A\hat{b} - 2A^T \hat{y} = 0$$
$$\hat{b} = \left(A^T A\right)^{-1} A^T \hat{y}$$

$$\text{When} \frac{\partial E}{\partial \hat{b}} = 0, \hat{b} \cong b \text{ and } \hat{y} \cong y \qquad \textbf{(3-12)}$$

$$b = \left(A^T A\right)^{-1} A^T y \qquad \textbf{(3-13)}$$

Where:

$$A = \begin{bmatrix} 1\, x_{11}\ x_{12} \ldots x_{1q} \\ 1\, x_{21}\ x_{22} \ldots x_{2q} \\ 1\, x_{p1}\ x_{i2} \ldots x_{pq} \end{bmatrix} \quad p: \text{number of data pairs}, q = \text{number of inputs} \qquad \textbf{(3-14)}$$

$$b = \begin{bmatrix} b_0 & b_1 & b_2 \dots b_q \end{bmatrix}^T \qquad \textbf{(3-15)}$$

$$y = \begin{bmatrix} y_1 & y_2 & y_3 \dots y_p \end{bmatrix}^T \qquad \textbf{(3-16)}$$

When a new data pair becomes available in the LS method all historical and new data are used to recalculate the unknown parameters, however for an online system, processing time is very important and utilizing this method is a time consuming process. Thus calculating the modified parameters using the new data pair and the previously estimated parameters is necessary without having to use the previous data. This is called a recursive least square method. The solution for a weighted version of the recursive least square method with a forgetting factor (λ) for time varying systems is described below:

$$b_w = \left(A^T W A\right)^{-1} A^T W y \qquad \textbf{(3-17)}$$

Assume $P_w = \left(A^T W A\right)^{-1}$ then

$$b_w = P_w A^T W y$$

$$W = \begin{bmatrix} w_1 & 0 & 0 \dots 0 \\ 0 & w_2 & 0 \dots 0 \\ 0 & 0 & w_3 \dots 0 \\ . & . & \dots\dots\dots \\ 0 & 0 & 0 \dots w_p \end{bmatrix}$$

The unknown parameters are updated iteratively by:

$$b_{k+1} = b_k + w_{k+1} P_{k+1} a_{k+1} (y_{k+1} - a_{k+1}{}^T b_k) \qquad \textbf{(3-18)}$$

$$P_{k+1} = \frac{1}{\lambda}\left[P_k - \frac{w_{k+1} P_k a_{k+1} a_{k+1}{}^T}{\lambda + a_{k+1}{}^T P_k a_{k+1}}\right] P_k \qquad \textbf{(3-19)}$$

λ is the forgetting factor used for time varying systems.

3.2.7 Evolving Clustering Method (ECM)

The process of dividing a big dataset into groups of similar objects that have similar properties within and dissimilar properties outside is called clustering. There are three main types of clustering: Hierarchal based clustering, Partitioning based cluster and density based clustering methods, a good survey for clustering can be reviewed in (Rai and Singh 2010). The Evolving Clustering Method (Song and Kasabov 2001b) is a distance based grid partitioning method that can be applied online due to its incremental nature through a one pass process however it can also be applied offline by including a constrained minimisation criteria. The distance measure can be Euclidean, Normalized Euclidean distance, equation 3-20 or any other appropriate distance measure. One important thing to make sure of when using this method is that the input data need to be normalized first. Otherwise, when using variables with different units, small/large distances might get confused due to the different inputs scales. A threshold value is defined at the beginning of the process which will decide the number of clusters to be created.

An example that shows how the ECM algorithm works online is presented in Figure 3-10, starting from the arrival of the first data pair P1 which is assumed to be the first cluster C1, then the second data pair P2 arrives and the distance between P2 and C1 is measured and compared to the threshold value Dthr, the distance is found to be less than Dthr, so the cluster C1 was moved to a new location C1' which is on the line connecting the two points P1 and P2. P3 arrives and the distance between this point and the existing cluster C1' is measured and found to be more than Dthr so a new cluster is added C2 with a zero radius and a position the same as the position of P3. Finally, P4 arrives and the distance between this point and existing clusters C1' and C2 is measured, P4 is found to be closer to cluster C1' with a distance less than Dthr, so the cluster centre is moved again to a new location C1'' along the line connecting C1' and P4. P1, P2 and P4 belong to cluster C1'' and P3 belongs to cluster C2. Steps 1 through 6 describe the process in more details:

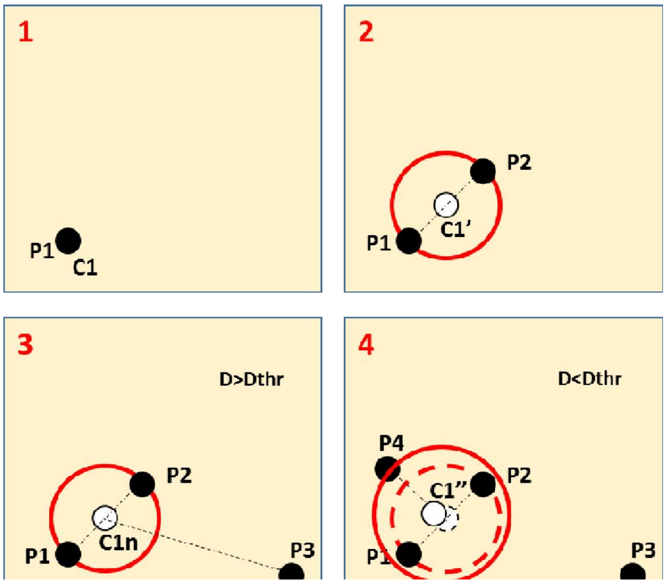


Figure 3-10 An Online ECM Process

__Step 1:__ Create the first cluster using the first data pair and assign a zero value to its radius.

__Step 2:__ When a new data pair becomes available, measure the distance between this data pair and all available clusters using the normalized Euclidean distance described in equation 3-20:

$$D(x,y)=\sqrt{\frac{\sum_{i=1,\dots,q}(x_i-y_i)^2}{q}} \tag{3-20}$$

__Step 3:__ Find the minimum distance and from which cluster.

__Step 4:__ If this minimum distance is less than the radius of this cluster then this data pair belongs to this cluster. In this case, none of the existing clusters will be updated, nor any new cluster is added.

__Step 5:__ If this minimum distance is more than the radius of this cluster, calculate a new distance as the distance from each cluster plus the radius of each cluster. If the new minimum distance is more than twice the threshold level, then a new cluster is created with its centre as the data pair position and its radius is zero.

Step 6: If the minimum new distance is less than twice the threshold level, then the radius of this cluster will be updated to the minimum new distance divided by 2 and the cluster centre is moved along the line connecting the new data pair and the old cluster centre. In this way the maximum distance from any data pair belonging to a cluster and its cluster centre is kept below the threshold defined at the beginning.

3.2.8 Particle Swarm Optimisation (PSO)

The idea of particle swarm optimization was inspired by the birds swarm and the way they behave and move towards their final destination (target). The method was first proposed in (Kennedy and Eberhart 1995) for optimisation of nonlinear functions and hundreds of papers over the last decade were published with certain modifications and hybrid utilization of this method with other artificial intelligence methods for optimisation and training of neural and neuro-fuzzy networks purposes. The swarm consists of a number of particles; each has an initially generated random position and velocity as it moves. Those particles will have three different options of movement:

1. Stay in their position.
2. Move towards their best found local position.
3. Move towards their best found global position.

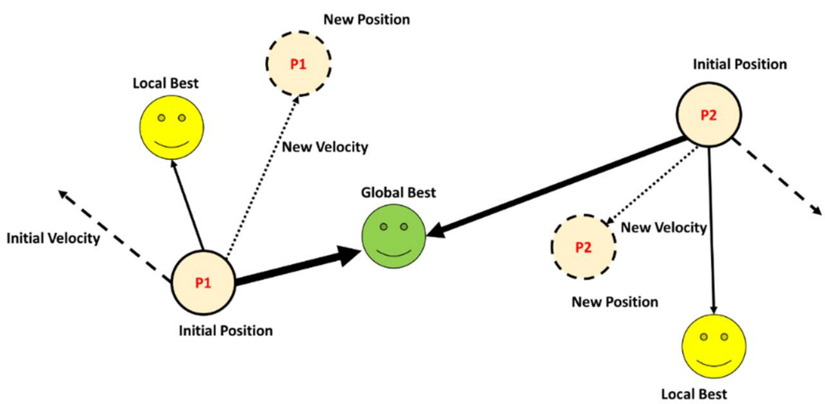


Figure 3-11 Particle Swarm Optimisation Example

Figure 3-11 shows two particles with their initial position (P1 & P2) and velocity. They cooperate with each other and know each ones local best and global best and have the option of either moving towards their local best or global best. Ultimately the final goal is to arrive at the global best position which represents the global best optimised solution.

Two basic equations 3-21 and 3-22 describe the movement of these particles and are used to update each particle's velocity and position:

$$v_i(t+1) = \omega v_i(t) + \psi_1 r_1 \left(Globalbest - x_i(t)\right) + \psi_2 r_2 \left(Localbest - x_i(t)\right) \quad \textbf{(3-21)}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad \textbf{(3-22)}$$

Where:

$v_i(t+1)$, $v_i(t)$:are velocity at time t+1 and t

$x_i(t+1), x_i(t)$:are the position at time t+1 and t

ω :Inertia weight

$\psi_1 r_1$, $\psi_2 r_2$, the first term of each is the global and local weights respectively, ideal value of 2 for each is used as a standard (Kennedy and Eberhart 1995) for no good reason apart from empiricism. The second term of each is a uniformly random number in the range [0, 1].

The Inertia weight can be a constant (Shi and Eberhart 2002), random (Eberhart and Shi 2002), adaptive (Nikabadi and Ebadzadeh 2008), sigmoid increasing (Mail et al. 2007), sigmoid decreasing (Mail et al. 2007), linear decreasing (Xin et al. 2009), chaotic (Feng et al. 2008), chaotic random (Feng et al. 2008), oscillating (Kentzoglanakis and Poole 2009), global-local best (Arumugam and Rao 2006), simulated annealing (Al-Hassan et al. 2007), natural exponent (Li and Gao 2009), logarithm decreasing (Chen et al. 2006) and exponent decreasing (Gao et al.

2008) inertia weights. The results of reviewing different inertia weight strategies have shown that based on accuracy and speed of convergence, the chaotic and random Inertia weights are the best used strategies respectively (Bansal et al. 2011).

The particles positions and velocities are randomly chosen at the start of the optimization process the particles positions and velocities are randomly created. A cost function or fitness is used to score each particle. The initial local best positions are the already created random positions, and the initial global best position is the particle position with the minimum fitness function. The positions are then updated using equations 3-21 and 3-22. The fitness function is used to score the newly created positions. The minimum fitness function is compared to the previously saved global best fitness, if this value is lower than the current global best fitness, the global best position will be substituted with this position else the global best position will remain unchanged. The fitness of each particle will be compared with the local best fitness of each particle; all particles with lower fitness will have new local bests as the new positions of those particles. The process will continue until the number of iterations is finished or a minimum value is achieved. The particles in the swarm cooperate with each other and share information about local bests and global bests, this information is used to adjust the particles velocity to move towards either the local best or global best.

3.2.9 Autoregressive Integrated Moving Average (ARIMA)

Merriam Webster Dictionary definition of a time series is: "*a set of data collected sequentially usually at fixed intervals of time*" (Merriam-Webster). Time series dataset is a set of data points collected periodically (at a regular period of time in seconds, minutes, hours...) that takes the following form, equation 3-23 (Shmueli 2011):

$$X = \{x_1, x_2, \ldots, x_n\} \text{ where } t = t_1, t_2, t_n \qquad \textbf{(3-23)}$$

i.e. it has temporal ordering which means that this data might have an internal arrangement like:

1. **A trend (upward or downward)**: long term change of the mean value of a time series
2. **Seasonal or Cyclic Behaviour**: a structure that repeats itself at certain period of time like the weather conditions.
3. **Autocorrelation:** autocorrelation of a signal is the cross correlation of this signal with itself. Applications like identifying a periodic signal masked by noise.
4. **Chaotic Behaviour**: Appears irregular or unpredictable however under certain initial conditions can actually be determinant governed by nonlinear deterministic laws.

Examples are: Process data collected using data historians and condition monitoring DAQs like: temperature, pressure, flow, valve opening position, motor current, voltage, vibration, etc. This data can be sampled at various sampling rates limited by the capabilities of the hardware/software used.

Time series forecasting is a step forward in the time series analysis field concerned with studying and predicting the time series future index or value based on past and present values.

In general, a time series has two distinct parts combined either in a multiplicative or additive way: systematic and non-systematic parts. The systematic part is a combination of all or any of the following: level (mean or average), trend (constant, linear, exponential, or other type) and seasonality (additive or multiplicative). The non-systematic part is Noise (measurement errors, some other reason) (Shmueli 2011).

Autoregressive Integrated Moving Average (ARIMA) is one of the statistical modelling methods that take into consideration the non-stationary effect in the time series normally by differencing the original signal to remove the non-stationary part. ARIMA is usually expressed in the following format: ARIMA (p, d,q). The three main parts of the name are a reflection of:

1. **AR(p):** Autoregressive of order p, indicating the number of autoregressive parameters used, this part of the model is based on autoregression.

2. **I (d):** Integrated of order d, indicating the number of differences needed to make the time series stationary, this process is called integrating.
3. **MA(q):** Moving Average of order q, indicating the number of moving average parameters used, this part of the model is based on the average of errors.

The 1st order autoregressive model AR (1) is the simplest autoregressive model. It is described by the following equation 3-24:

$$y_t = a_1 y_{t-1} + e_t \tag{3-24}$$

Where y_t and y_{t-1} are the current and previous values. e_t is the noise term sometimes called random shock. a_1 is the unknown parameter to be identified.

For higher order autoregressive models, more previous values (lags) of y_t are included. The autoregressive model is called autoregressive because the regression is done on the time series itself i.e. the output of regression is a function of its previous values not some other independent time series.

The 1st order moving average model MA (1) is the simplest Moving average model. It is described by the following equation 3-25:

$$y_t = e_t + c_1 e_{t-1} \tag{3-25}$$

e_t and e_{t-1} are the current and previous values of the random shock (Noise). c_1 is the unknown parameter to be identified. The MA model includes previous (lagged) terms of the random shock (noise).

The following methodology for building ARIMA (p, d, q) model is proposed in (Shieh and Hung 2009):

1. **Step 1: Identification** of the Model using autocorrelation and partial correlation plots.
2. **Step 2: Estimation** of the Model unknown parameters.
3. **Step 3: Evaluating** the identified Model(s) ensuring random model error and that the parameters identified are statistically significant.

4. **Step 4:** Select the most accurate Model and use for **forecasting**

A detailed explanation of the above steps is described below:

Step 1- Identification: Testing if the time series is stationary or not is the 1st step in identifying the model's order. This can be done either visually using an XY plot or using a stationary testing method like Dickey Fuller Test. The main characteristics of a stationary time signal is that it needs to have a constant mean, constant variance and covariance and correlation between y_t and y_{t-n} is the same at all times. A number of differencing steps might be required to make the time series stationary.

Autocorrelation and Partial correlation plots are bar charts showing the correlation and partial correlation coefficients of the time series and its lags. Significant lags are lags that exceed the 95% confidence band in the plot.

Autocorrelation and Partial correlation are calculated using the below equations 3-26 and 3-27:

$$Autocorrelation = \frac{Covariance(y_t, y_{t-n})}{Variance(y_t)} \quad \textbf{(3-26)}$$

$$Partial\ Correlation = \frac{Covariance(y_t, y_{t-n}|y_{t-n+1}, \dots, y_{t-1})}{Variance(y_t|y_{t-n+1}, \dots, y_{t-1})Variance\left(y_{t-n}\middle| y_{t-n+1}, \dots, y_{t-1}\right)} \quad \textbf{(3-27)}$$

Some rules of thumb that helps identifying the order of the ARIMA model are:

1- If the autocorrelation plot dies slowly over the different lags or fluctuates around zero but dies slowly to zero this means more AR terms are required
2- If the autocorrelation plot shuts off abruptly this means more MA terms are required
3- From 2, the number of significant lags before the plot shuts off abruptly define the order of the MA
4- The order of AR is best defined using a partial correlation plot
5- The order of AR is the number of non-zero lags in the partial correlation plot before which the plot shuts off.
6- The order of MA is best defined using the autocorrelation plot

7- The partial correlation plot for MA models will die slowly over the different lags. However, the autocorrelation plot will give better representation of the order of the MA model which is equal to the number of non-zero lags before the plot shuts off.

8- If both the autocorrelation and partial correlation plots don't tail off over many lags (amplitude stays close to 1), the time series isn't stationary.

Examples autocorrelation and partial correlation plots are shown in Figure 3-12.

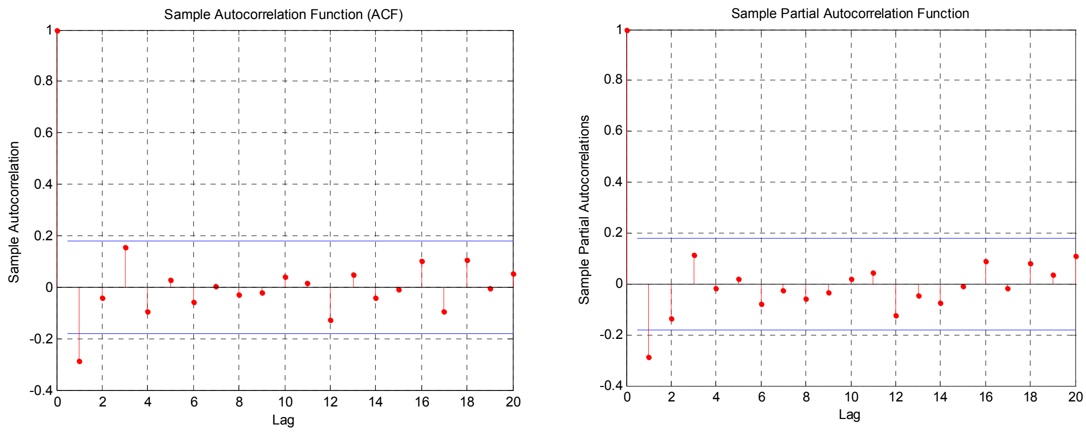


Figure 3-12 Auto-correlation and Partial Correlation Plots Example

Step 2- Estimation: It is an easy job to estimate the parameters for the autoregressive model using least squares method, however, it is more complicated when it comes to the integrated and moving average models. It's usually a best practice to utilise one of the commercial applications to do this job. In fact, some Softwares have the capability of also doing step 1 above through an iterative way to find out the best ARIMA model and its parameters.

Step 3- Evaluation: One of the expectations from the ARIMA model is that the residual is random i.e. the autocorrelation is equal to zero or at least fluctuates around zero within the 95% confidence band in the plot. The residual is expected to be normally distributed around zero (mean value) with a constant variance. A histogram plot and normal Q-Q plot can be used to test that. The second element of evaluation is related to the estimated parameters and whether they are statistically significant or not i.e. they should be significantly more than zero. Below is an example of how this can be tested:

The variance of the estimated parameters for an autoregressive model, for example for a_1 is, equation 3-28:

$$Var = \frac{\left(1 - a^2{}_1\right)}{N}, \text{ Where } N \text{ is the time series samples number} \tag{3-28}$$

With a 95% confidence band equal to two standard deviations around the estimated parameter a_1, equation 3-29:

$$95\% \text{ Confidence Band} = a_1 \pm \sqrt{Var} \tag{3-29}$$

The parameter estimated is considered significant if the 95% confidence band doesn't include zero and infact it should be well away from zero. (Coats et al. 2009) Includes similar equations for MA and ARMA models.

Step 4- Forecasting: After the appropriate ARIMA model is selected, the time series future values can be predicted.

ARIMA models are commonly used in so many different areas like: economic and business applications, statistics, machine learning and weather forecasting to name a few. ARIMA was compared by several researchers to ANFIS in different applications showing that ANFIS outperforms this method (Fahimifard et al. 2009), (Yayar et al. 2011) and (Cesar et al. 2009). However, since in most of the cases the future inputs to the NeuroFuzzy model aren't available, they need to be predicted and then used in this model. ARIMA can be used for such purpose.

3.2.10 Empirical Mode Decomposition (EMD) Method

Assuming that the signals are stationary and linear is the basis for time and frequency domain traditional signal processing methods like FFT (Fast Fourier Transformation). Hence, signals that aren't stationary or linear will introduce various signal processing errors when an algorithm like FFT is applied. Another issue is related to arriving at a compromise between a good resolution in the time and frequency domains. To deal with signals that are non-stationary and/or nonlinear, there is a need to tackle these issues by introducing new signal processing techniques which can deal with these issues. Empirical Mode Decomposition (EMD) is first introduced in (Huang et al. 1998) as an adaptive signal processing method that deals with non-stationary and nonlinear signals in the time domain away from the frequency domain.

This method attracted many researchers in different fields, examples including: financial data (Hong 2011), Medicine and Biology (Han-Oh et al. 2012) and (Balocchi et al. 2004), Climate and Earthquake (Huang et al. 2001), (Barnhart and Eichinger 2011) and Fault diagnosis of mechanical systems (Yu et al. 2005), (Bin et al. 2012), (Dybala and Zimroz 2014) and (Lei et al. 2013), etc.

EMD is the first step in a Hilbert Transform algorithm. The basic idea of it is that the signal is decomposed into a residual signal and a set of what is called IMFs (Intrinsically Mode Functions) - simple oscillatory components satisfying the following conditions (Huang et al. 1998):

1. The number of maxima and minima points and the number of zero crossing points in the whole dataset need to be equivalent or more by a maximum of one point.
2. The envelope mean of the upper and lower envelopes (defined by the local maxima and local minima) is zero.

Figure 3-13 describes the end to end EMD process.

The following Mackey-Glass time series, equation 3-30, which is developed to model the white blood cells reproduction is used as a benchmark example by hundreds of researchers in the literature and will be used here to give an example for the EMD process.

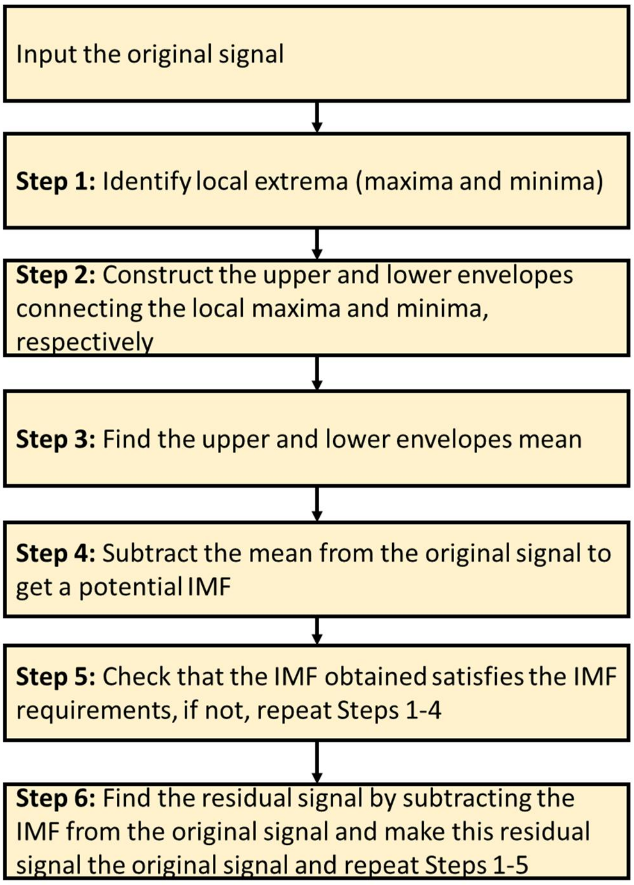


Figure 3-13 End to End Process of EMD

$$\frac{\mathrm{dx(t)}}{\mathrm{dt}} = \frac{0.2x(t-tau)}{(1+x(t-tau)^{10})} - 0.1x(t) \tag{3-30}$$

Where:

$x(t)$: is the density of mature cells in blood circulation

tau : is the delay constant with a given value of 17

$x(0)$: is the initial condition equal to 1.2

The above Mackey-Glass differential equation was solved using the 4th order Runge-Kutta numerical method in Matlab, Figure 3-14.

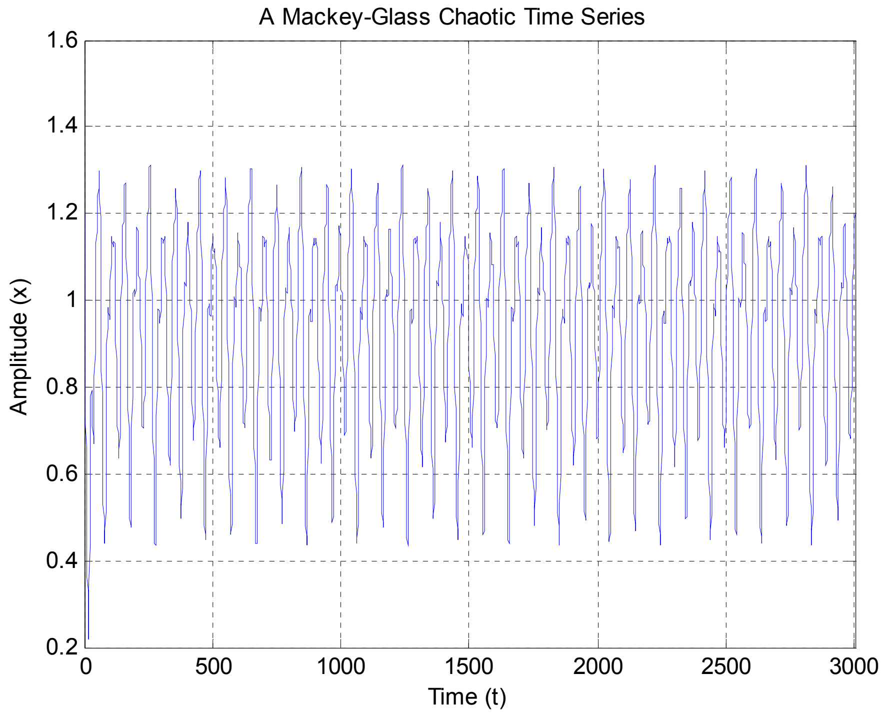


Figure 3-14 Mackey Glass Time Series Plot

The first 200 data points are removed from the signal and only the 800 consecutive points are utilised in the EMD simulation. Figure 3-15 shows the results of simulation for the original signal (first trend), the residual signal (second trend) and the first three IMFs named x, y and z in Figure 3-15. It's expected that the addition of all IMFs and the residual signal is equal to the original signal. To test this hypothesis, Figure 3-16 shows the original signal superimposed on the sum of all IMFs plus the residual signal, the error between the two is show in Figure 3-16 with a mean absolute error of 7.5495e-17 which is pretty good considering this approximation method.

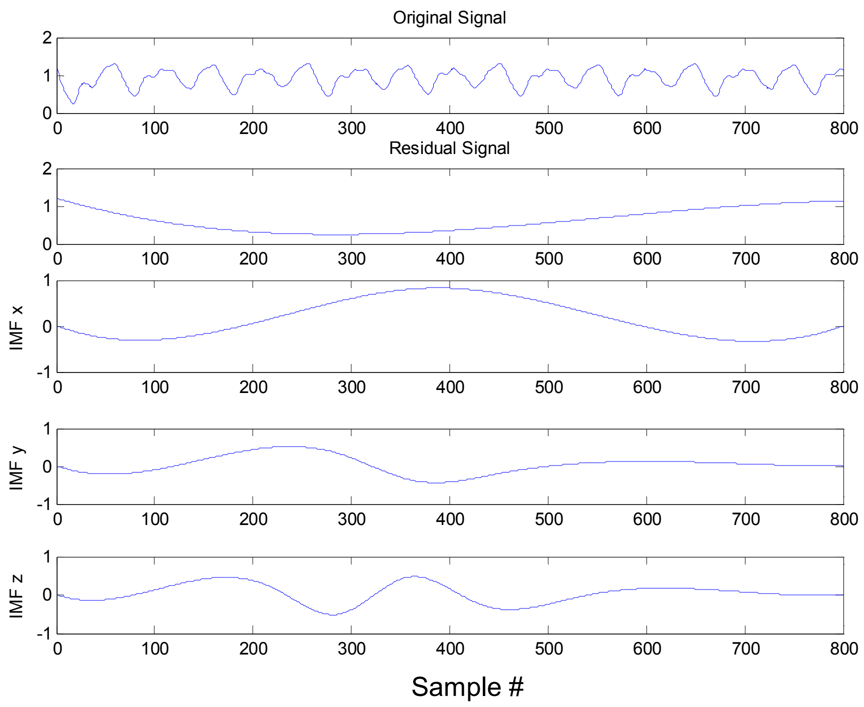


Figure 3-15: EMD for Mackey-Glass Time Series

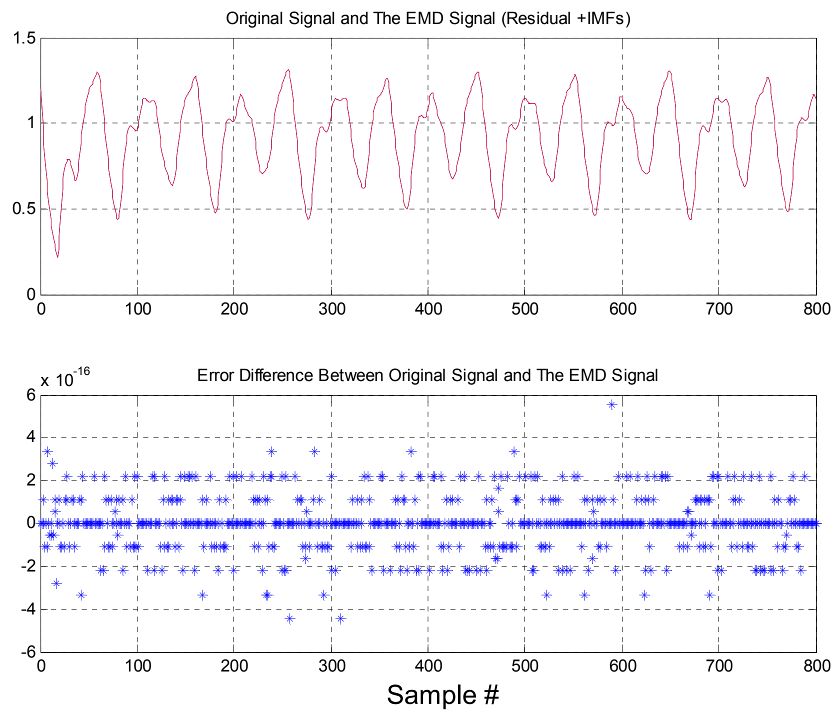


Figure 3-16 Mackey-Glass Time Series Superimposed on the EMD output and the Error Difference

Several issues with the EMD method are highlighted in the literature as described below (Lei et al. 2013):

1. End effects: Distortions at both IMF sides caused by the EMD model itself.
2. The IMFs with each other aren't strictly orthogonal
3. A problem of mixing between modes sometimes happens.
4. Lacks a theoretical foundation
5. The first IMF covers a large frequency range.
6. Interpolation Problems
7. Stopping Criterion
8. Selection of best IMFs for inclusion in the analysis

Several researches in this area were conducted to address the above issues. New algorithms and a hybrid between several algorithms are used. Examples of the new algorithms, can be found in (Lei et al. 2013) specifically for fault diagnosis of rotating equipment: Ensemble Empirical Model Decomposition (EEMD), Wavelet Packet Transform- Empirical Mode Decomposition (WPT-EMD), B-Spline Empirical Mode Decomposition (BS-EMD), methods combining EMD with other methods like independent component analysis, Wigner-Ville distribution, order tracking, wavelets,etc. The original EMD algorithm will be used here with potential future integration of some of the above mentioned modifications for comparison purposes.

3.2.11 Statistical Process Control: Tracking Signal

Tracking signal is one of the techniques used by researchers for uncertainty management and process quality control applications. Tracking Signal was first introduced in (Brown 1959) as the ratio between the cumulative forecasting error and the mean absolute deviation (MAD) as per equation 3-31:

$$Tracking\ \ Signal\ = \frac{Cumulative\ \ Error}{MAD} \tag{3-31}$$

The mean absolute deviation is calculated by adding the absolute forecasting error in each period and taking the average of this total. From the process quality control theory, this value is continuously compared to the control limits defined by the user. While the tracking signal value is within the control limits the prediction process is in control. A value of ±5 is used here in this research however from reviewing the literature a value between ±2 and ±5 is normally used. This is normally plotted on XY plot with the X axis representing the number of prediction steps and the Y axis representing the tracking signal value as shown in Figure 3-17.

This research will use the tracking signal method to detect any deviation from the normal operating envelope of machinery for the first time according to the author's knowledge.

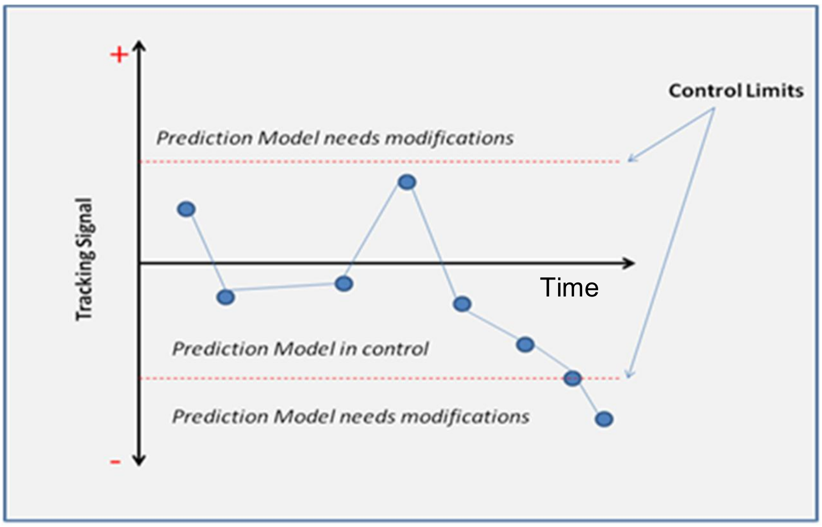


Figure 3-17 Tracking Signal Example

3.3 Architecture of the Proposed Condition Monitoring and Prognostics Framework

3.3.1 Introduction

Section 3.2 described in details all the enabling techniques that will be utilised in the proposed condition monitoring and prognostic framework. This section will describe where each one of these methods is used within the proposed framework and how they all work together. The proposed Framework is fully aligned with ISO 17359:2011 (BS ISO 17359:2011 2011) and is described in more details in section 3.3.2. The proposed framework is shown in Figure 3-18 in a generic format as an end to end process for condition monitoring and prognostic. The main steps of this framework are listed below:

- **Step 1:** Data Collection and Validation
- **Step 2:** Data Processing and Feature Extraction
- **Step 3:** Fault Classification (Diagnostics)
- **Step 4:** Fault Progression Prediction (Prognostics)
- **Step 5:** Maintenance Action
- **Step 6:** Continuous Improvement

The above steps will be discussed in more details below however its worth mentioning that this is assuming that the list of equipment being monitored is selected based on their criticality rating and the parameters to be monitored are selected based on a FMECA/FMEA type study linking failure modes to the parameters sensitive to detecting these failures. The literature survey chapter 2 covers a review of the current practices, models and algorithms related to condition monitoring, diagnostics and prognostics as such this framework is intended as a generic one with more emphasis on the author's novelty integrated condition monitoring and prognostics framework.

Plenty of work has been done so far exploring the first step, data collection and validation, as such this will not be described any further in this document, however one thing that is worth mentioning is that it is very important to ensure noisy data is removed before proceeding with any offline type modelling and a

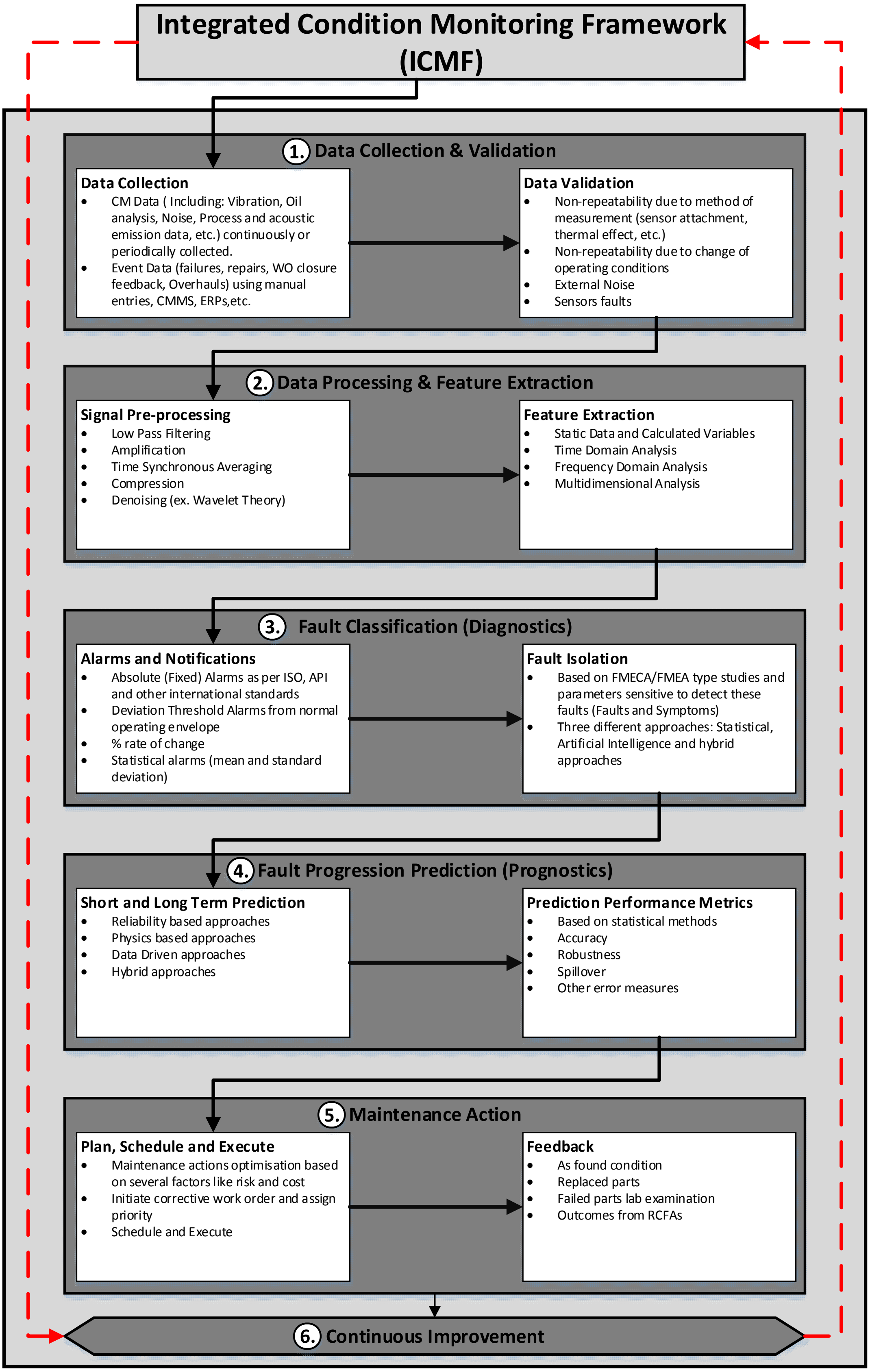


Figure 3-18 Integrated CM and Prognostics Framework

capability of the system to detect any noise, outliers, NaNs, etc. is in place where online modelling and processing is required as this will impact the quality and confidence of the results.

3.3.2 Integrated Condition Monitoring and Prognostics Framework

The Integrated Condition Monitoring and Prognostics Framework proposed in this work is aligned with ISO 17359:2011 (BS ISO 17359:2011 2011) and is based around the I-P and P-F curve, Figure 3-19 and the location of the health indicator (condition monitoring parameter) along this curve as function of time. This curve describes the machine condition as observed by a sensitive health indicator. During normal working conditions, called the normal operating envelope which is the green area of the I-P and P-F curve between the points I (Installation New Condition) and P (Potential Failure Detection). Fluctuations around some mean value can occur as a result of variations in the operating and environmental conditions. However, these fluctuations shouldn't exceed the upper and lower limits of the normal operating envelope for this health indicator.

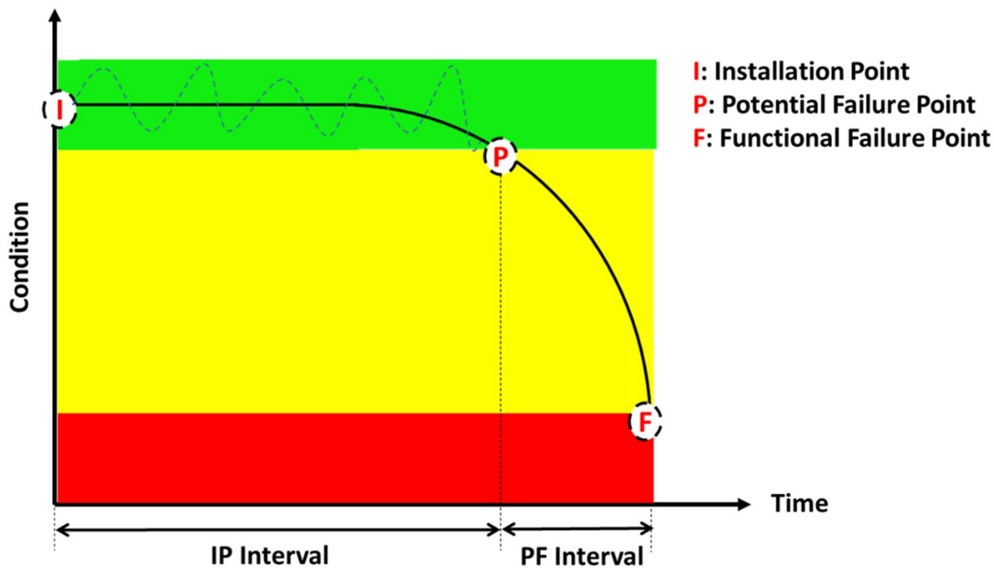


Figure 3-19 IP and PF Curve

As the machine health deteriorates and a fault starts developing over time the health indicator(s) sensitive to that specific failure mode will start changing (increasing or decreasing) as an indication of a developing fault. By the time, the health indicator is at the P point (Potential Failure Detection), a potential failure is imminent and the way how the progression of failure rate and shape will be, will rely on so many different covariates including the operating and environmental conditions for example, this might even initiate a development of a secondary and tertiary dependent failure modes. As such an online process to monitor the fault progression over time (short term and long term) with capabilities of detection of any transient changes in the trends is very important. Fault Isolation is normally done within the P-F interval (The yellow region), F (is the Functional Failure point). Once the fault is isolated and an estimate for the remaining useful life is made, maintenance work can be planned, scheduled, materials is ordered (if required) and the work is executed and closed with useful feedback about the as found condition, any replaced parts, failed parts lab examination and outcomes from the RCFA (if this was done).

The maintenance work feedback is very important as this will aid in the continuous improvement cycle for the condition monitoring and prognostics program. Depending on the frequency of taking measurements, some failure modes might go undetected as they have very short P-F intervals less than the frequency of taking the measurements. Where failure modes with short P-F interval are highly likely to occur with expensive consequence of failure (Safety Related/ Environmental and/ or financial), an online system should be in place or the frequency of taking offline measurements should increase to becomes at least 1/3 the P-F interval as a rule of thumb, this rule of thumb is only a recommendation from experience but has no fixed background to refer to.

Although only 11% of all failures are age related (the remaining 89% are random) and can be trended and monitored over time, most of the random failures exhibit some pre-warning of some sort when they are about to develop. Different health indicators will have different P-F curves depending on the sensitivity of those indicators in detecting the specific failure mode. Some are more sensitive than others, some aren't sensitive at all for that specific failure mode.

Assuming that the data is collected using some sort of data historian or condition monitoring system either online or offline, Figure 3-20 shows the remaining steps of the novelty integrated condition monitoring and prognostics framework proposed with all enabling techniques as described in section 3.2 and their role in the process. The main core of this process as mentioned before is the I-P and P-F curve. The following subsections will describe each step of the process in more details:

1- Normal Operating Envelope Monitoring (within I-P Interval)
2- Fault Classification and Diagnostics (within P-F Interval)
3- Short Term Fault Progression Prediction (within P-F Interval)
4- Long Term Fault Progression Prediction (within P-F Interval)

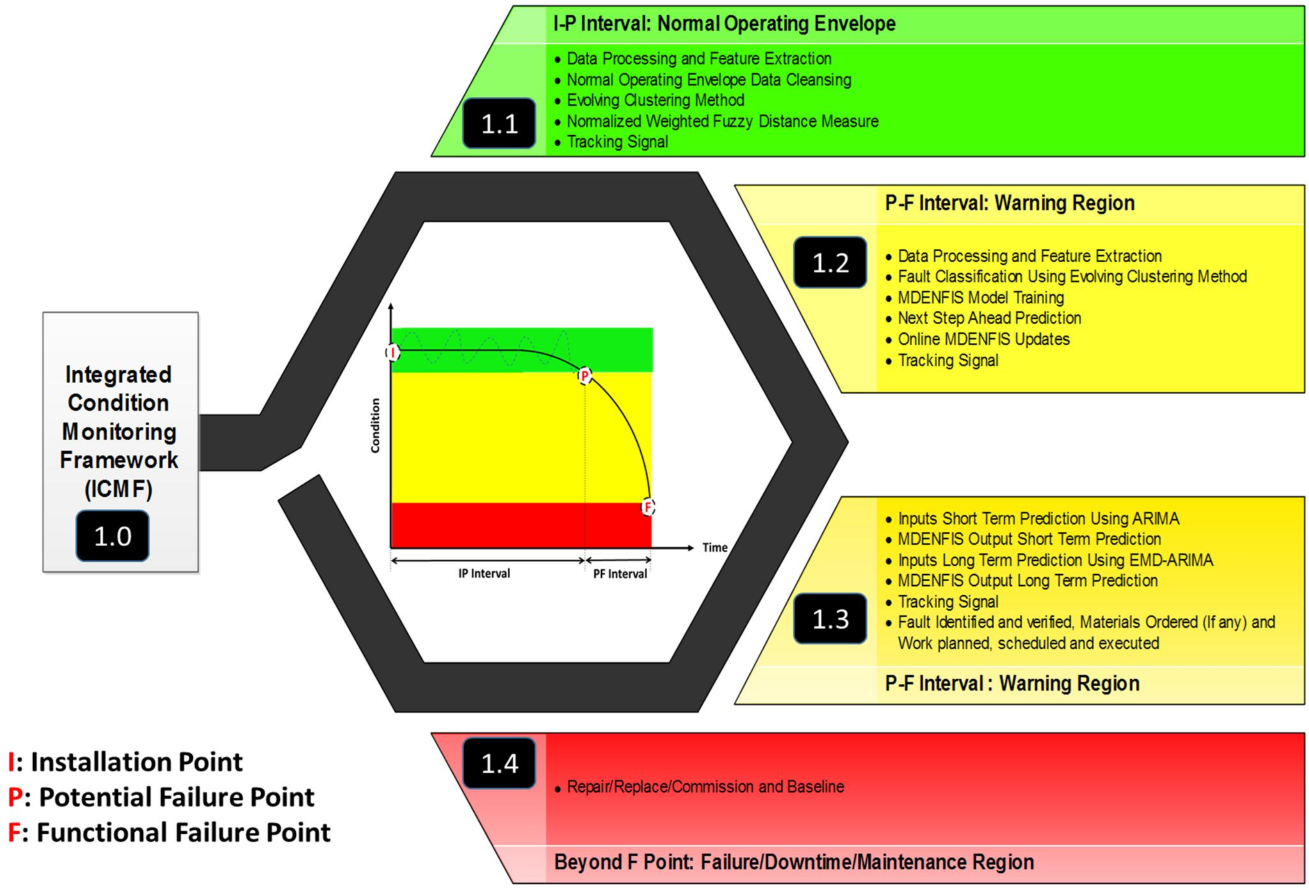


Figure 3-20 Proposed Integrated Condition Monitoring Framework (ICMF)

The P-F interval can be easily determined if the failure mode is age related and has a linear progression curve, however it gets really difficult to determine this interval when other random failure modes are involved (BS ISO 17359:2011 2011) or even when more than one failure mode occur at the same time. Subject

Matter Experts (SME) who have long experience with similar types of machines might give some estimate about this interval however there is a high uncertainty in these estimates as there is no fixed background for them and depending on the operating conditions and their effect on the progression rate the same failure mode might have more than one interval. This research will try to estimate this interval using data driven approaches.

3.3.3 Normal Operating Envelope Monitoring

Normal Operating Envelope (NOE) is a zone defined by an upper and lower limits within which adjustments to the operating conditions of the machine will have no adverse effects on the integrity of the machine. Drifting outside this zone will put the integrity of the machine at risk.

NOE can be defined for one or more parameters and can be viewed easily in 1-dimensional, 2-dimensional and 3- dimensional graphs for 1, 2 and 3 parameters respectively. Special Softwares will be required to view the NOE for more than 3 parameters at the same time.

NOE is normally expressed in terms of pressure, temperature, flow, speed, etc. i.e. process parameters however this definition can be extended to cover other health indicators, like for example vibration and oil analysis. The first reference for defining the NOE is the Original Equipment Manufacturer (OEM) recommendations. This can also be done from experience with the same or similar machinery by SMEs. For condition monitoring data, in addition to the above; limits can be calculated from the past running history of the machine using statistical methods or, ISO/API defined limits can be utilised instead.

It is very important when defining the NOE to take into consideration the interaction between the operating and environmental conditions with other process and health indicators.

The proposed methodology for monitoring where the machine is running in relation to its NOE and detecting at an early stage any drifting outside this zone to allow reversing the effect of that behaviour with no or minor consequences is explained in Figure 3-21.

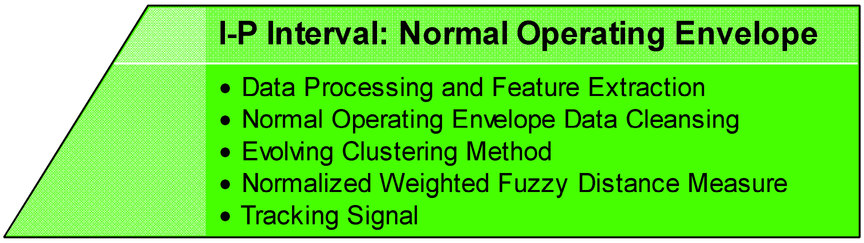


Figure 3-21 Normal Operating Envelope Region

The following enabling techniques will be used here:

1- Evolving Clustering Method
2- Normalised Weighted Fuzzy Distance Measure
3- Tracking Signal

Once the data acquired is processed and the relevant features and health indicators are extracted, the data is cleansed to only contain any variations within the NOE, this means that any zero speed/amps data indicating that the machine isn't running, bad tags, and abnormal behaviour outside the NOE will be removed. This can be visually done using one of the visualisation Softwares.

ECM is then used to reduce the number of operating states by removing any similarities between the different states due to repeated measurements at the same operating state. Now that the operating states have been defined, those are the cluster centres obtained by applying the ECM algorithm, the model is ready to work online. As new data arrives, the normalised fuzzy weighted distance is calculated by fuzzifying the input using rectangular membership functions and the cluster centres, other membership functions like the Gaussian curve, triangular and sigmoid membership functions can be also be used. The idea is that using a hybrid of ECM and normalised fuzzy weighted distance measure, the ideal behaviour representing where the machine is expected to be running will be estimated and compared with where the machine is actually running through a tracking signal method.

The ideal behaviour estimate is obtained using the below steps:

Step-1: Use ECM to obtain the NOE matrix which contains what is called the clusters centres. The model is now ready to work online.

Step-2: This step is used to map crisp input data (new data) to fuzzy outputs using the following triangular membership function, equation 3-32:

$$\mu_X(x) = \frac{1}{1+\left|\frac{x-c_i}{a_i}\right|^{2b_i}} \qquad \textbf{(3-32)}$$

x is the input values

a_i, b_i, c_i are the premise parameters used to define the fuzzy region

Step-3: The output of this step is the product of the fuzzified output of each input variable to the same cluster and for all clusters. This will calculate the firing weight from each cluster, equation 3-33.

$$w_i = \mu_{X_i}(x).\mu_{Y_i}(y).\mu_{Z_i}(z) \qquad \textbf{(3-33)}$$

Step-4: This step is used to calculate the normalised distance weight from each cluster, equation 3-34.

$$\overline{w}_i = \frac{w_i}{w_1+w_2} \text{ where } i = 1,2 \qquad \textbf{(3-34)}$$

Step-5: The output of this step is the estimate value of the ideal behaviour, equation 3-35.

$$O = \sum \overline{w}_i.c_i \qquad \textbf{(3-35)}$$

Step-6: The tracking signal method is then used to compare the actual data coming in with the ideal estimate in relation to the NOE. Any drifting outside the NOE zone is detected and flagged. Examples using benchmark and field data for this method are included in Chapters 4 and 5.

3.3.4 Fault Classification and Diagnostics

A clear definition and distinction between a failure and a fault is essential before discussing fault classification and diagnostics. ISO 13372:2012 (BS ISO 13372:2012 2012) differentiates between a failure and a fault through the following definitions:

- **Failure (An event):** "*Termination of the ability of an item to perform a required function*" (BS ISO 13372:2012 2012).
- **Fault (A state):** "*Condition of a machine that occurs when one of its components or assemblies degrades or exhibits abnormal behaviour, which may lead to the failure of the machine*" (BS ISO 13372:2012 2012). A fault can exist without a failure however a failure is always caused by a fault.

The definition of failure above is also the definition of a functional failure as seen in the I-P and P-F curve, Figure 3-19, as it is linked to the equipment inability to perform a required task to an acceptable standard. This definition covers both total failure which is the complete loss of function and partial failure which is the partial loss of function where the machine is still functioning but outside the acceptable zone (Moubray 1997).

Diagnostics is defined as: "*The examination of symptoms and syndromes to determine the nature of faults or failures*" (BS ISO 13372:2012 2012). Within the fault classification and diagnostics area we are more interested in diagnosing faults rather than failures, however, in certain situations where failures happen, classification and diagnostics will be working in a post mortem phase with the objective of establishing lessons learned and identifying early warning indicators that will help improving the detectability of these failures at an early stage in the future.

ISO 17359:2011 (BS ISO 17359:2011 2011) and ISO 13372:2012 (BS ISO 13372:2012 2012) have specified certain prerequisites for a robust diagnostics process, including the following:

1- Establish a FMEA or FMECA study as per IEC 60812 (IEC 60812 2006).
2- The following outcomes from 1 are required:
    a. List of major components and their functions
    b. List of failure modes and their causes on the component level
    c. A prioritised list of failure modes based on their criticality for inclusion within the condition monitoring program for a specific machine
    d. Failure Modes versus symptoms and health indicators matrix
3- Select measurement techniques, hardware and software from which the health indicators in point 2.d are computed or derived.

Figure 3-22 shows an example of failure modes versus health indicators matrix. The first two columns are the list of maintainable items (components) and failure modes. The third column including a list of all possible health indicators (measured directly or derived). For example a radial bearing (component) damage due to lack of lubrication (failure mode) can be detected using the following health indicators (parameters): Bearing metal temperature, bearing vibration and lube oil level.

The failure modes versus health indicators matrix aids as one of the inputs to an automated classification and diagnostics system. The proposed automated classification and diagnostics system is built based on the ECM algorithm for classification. The ECM is an unsupervised classification method however, users have the ability to include additional classes from experience (supervised classification). This will make the proposed classification and diagnostics system a semi-supervised classification method although it can work in unsupervised classification mode. The method will initially work in an offline mode for learning purposes by including classifications from historical data and additional classes added by the experts. The method will then work in an online mode to identify any fault condition. This method has also the capability of adding additional classes (failure modes) that weren't included during the offline learning phase due to lack of historical examples or being missed by the user. The output of this system is one of the following:

1- Normal Mode

2- Fault Mode (with a specific failure mode type)

3- New Mode (with unknown failure mode type).

The third output is a result of the machine operating in a new mode. This will be flagged to the user for further investigation. Findings of this investigation are fed back to the system by either assigning a specific failure mode type to this new mode or considering this mode as a normal one.

		Polytropic Efficiency (%)	Bearing Temp. (Deg C)	Bearing Vib. (Micron p-p)	Suction Pressure (bar a)	Suction Temp. (Deg C(	Discharge Pressure (bar a)	Discharge Temp. (Deg C)	Flow Rate (m3/hr)	Recycle Valve Position (%)	Lube Oil Level	Lube Oil Temp. (Deg C)	Lube Oil Pressure (bara)	Lube Oil Filter DP (bara)	Rotor Axial Position	Seal Gas Pressure (bara)	Seal Gas Filter Dp (bara)	Seal Gas Flow	Lube Oil Analysis
Maintainable Item	**Failure Mode Description**	**Matrix mapping relationships between Failure Modes and Observed Effects**																	
Radial Bearing	Damage due to lack of Lubrication		●	●							●								
Radial Bearing	Damage due to (brinelling, fretting wear etc.)		●	●															●
Thrust Bearing	Damage due to excessive vibration		●	●											●				
Rotor	Internals worn / damaged (impellor, O rings etc.)	●		●			●	●	●										
Rotor	Deposits buildup/Fouling	●	●	●			●	●	●										
Rotor	Loose parts			●															
Antisurge System	Surging						●		●	●					●				
Pressure Transmitter	Abnormal instrument reading (suction)				●														
Filter	Lube Oil Filter Blockage												●	●					
Dry Gas Seal	Failure to contain fluid															●	●	●	
Valve	Lube Oil PCV malfunction												●						
Piping and Support	Piping strain		●	●															
Baseframe	Soft foot		●	●															

Figure 3-22 Failure Modes versus Health Indicators Matrix

3.3.5 Short Term Fault Progression Prediction

Once the behaviour of the machine represented by the condition monitoring parameters (health indicators) started to drift outside the NOE, it is important to predict the machine future behaviour and the progression of the fault identified in the fault classification and diagnostics step to enable planning the maintenance task required in terms of time, scope of work, resources and materials, however this step can also be done even if no fault is identified at present. This prediction is normally done using historical and current values of the condition monitoring parameters (health indicators) to predict future values.

Two types of prediction models are available: Univariate and multivariate. The first one indicates that the future predictions of the variable of interest are based on past and current values of the same variable. The second one indicates that the future predictions of the variable of interest are based on a number of variables which might include past and current values of the variable of interest also. One important condition for selecting the input variables is that there need to be a strong correlation between each one of them and the variable of interest (the output) and a weak correlation in-between them to prevent redundancy and over fitting problems in the prediction model. To ensure the previous criteria is met, Principle Component Analysis (PCA) is applied for example.

Prediction models can be classified in a different way based on the number of outputs into multi-inputs single output (MISO) and multi-inputs multi-outputs (MIMO) models. Yet a third way of classifying prediction models is based on the length of the prediction horizon into short term prediction models including single step ahead prediction and long term prediction models. Both short term and long term predictions will be discussed as part of the proposed integrated condition monitoring and prognostics framework.

Figure 3-23 shows three main steps for the short term prediction:

1- MDENFIS Model Training
2- Next Step Ahead Prediction
3- Online MDENFIS Updates

Before discussing the proposed hybrid model MDENFIS (Modified Dynamic Evolving Neuro-Fuzzy Inference System) it is imperative that the origins of this model are covered.

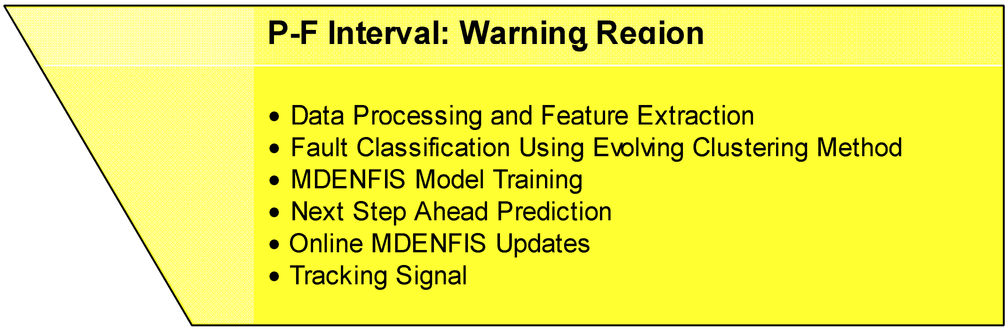


Figure 3-23 Warning Region Short Term Prediction

DENFIS was proposed in (Kasabov and Song 2002). This method integrates both Neural Network and Fuzzy Logic into a hybrid more powerful approach taking advantage of the Adaptive learning capability and reasoning capability of Neural Networks and fuzzy logic, respectively. The DENFIS system learns the Membership Functions parameters and rules incrementally purely from data using ECM (Song and Kasabov 2001b) which is a type of grid partitioning method based on a distance measure explained in details in section 3.2.7, the distance measure used here is the Normalised Euclidean distance. The FIS utilised is a 1st order Takagi-Sugeno-Kang (TSK) FIS. As a standard the triangular type MF is utilised in the model however in this work an exponential type MF is used instead. Through a comparison study between using the triangular MF and the exponential MF, the prediction accuracy is found better using the exponential MF hence using it in this work. The consequent parameters are learned using a weighted recursive least square method.

In the ANFIS model described before in section 3.2.7 the number of IF-THEN rules generated is 2^q rules where q is the number of inputs. When a large number of variables (inputs) is selected with the grid partitioning method, the curse of dimensionality will occur i.e. if for example the number of variables is 15 then the number of IF-THEN rules generated will be 2^{15}=32768 rules this is computationally very expensive and will take a very long time leading to an out of memory error however in the case of DENFIS the number of rules is equivalent to the number of clusters found by the online ECM. There are some similarity between the TSK model for ANFIS and DENFIS in terms of the main layers however the DENFIS structure is flexible allowing more new features and new rules to be added to the model incrementally over time. The DENFIS model consists of 5 main layers:

Layer 1: This layer contains adaptive nodes that have functions like the triangular and exponential membership functions which are used to map crisp input values to fuzzy linguistic expressions and vice versa. Other membership functions can also be used. Equation 3-36 shows the exponential membership functions used in this work.

$$\mu\ (x) = \exp\left(\frac{-(x-c)^2}{2\sigma^2}\right) \tag{3-36}$$

c : is the cluster center

σ : is the width

The width is usually used as the radius of the cluster however since the learning is done online, when a new cluster is added, initially a zero value is assigned to its radius hence the radius can't be used to represent the width of the cluster using this method without having some constraints or a value different than zero is used. However, a constant value can be used for all membership functions usually a fraction of the threshold value, for example. 0.9*Dthr.

Layer 2: Only fixed nodes are available in this layer. The output is the product of all input data to these nodes, equation 3-37.

$$w_i = \mu_i(x_1).\mu_i(x_2)...\mu_i(x_q) \text{ where } \ i = 1,2 \quad i^{th} \text{ rule firing strength} \tag{3-37}$$

Layer 3: Only fixed nodes are available in this layer. The output is the ratio between each i^{th} rule firing strength and the sum of all firing strength (normalization), equation 3-38.

$$\overline{w}_i = \frac{w_i}{w_1 + w_2} \text{ where } i = \text{cluster \#} \tag{3-38}$$

Layer 4: This layer contains adaptive nodes. The output from this layer is the product of the ith normalized firing strength and a polynomial function that has the consequent parameters, equation 3-39.

$$y_{4,i} = \frac{w_i . y_i}{\sum w_i} = \overline{w}_i .(b_1 x_1 + b_2 x_2 + ... + b_q x_q + b_0) \qquad \textbf{(3-39)}$$

where
$i = \text{c}luster\ \#$
$\text{q} = \#\ \text{of inputs}$

Layer 5: Only one fixed node is available in this layer. The output from this node is the sum of all inputs to this node and this will be the estimated output from this model which will later be compared to the actual output in order to evaluate the prediction accuracy, equation 3-40.

$$Output = y_{5,1} = \frac{\Sigma w_i . y_{4,i}}{\Sigma w_i} \qquad \textbf{(3-40)}$$

The MDENFIS novel model proposed here, Figure 3-24 is a DENFIS model modified by utilising a global optimisation method, PSO to optimise the solution obtained by the ECM method. To the authors knowledge this is the first time to use this hybrid learning approach for optimisation of the antecedent parameters.

The learning process starts with half the training data (or any other appropriate number of training input/output pairs) to obtain the antecedent parameters initial values. Whereas the consequent parameters are calculated using a weighted recursive least square method. The modified PSO is used to optimise the solution obtained by the ECM. Because the initial swarm in the standard PSO method is generated randomly it will normally take long time to converge, to address this issue an effective local approximation method (ELAM) is used, Figure 3-25.

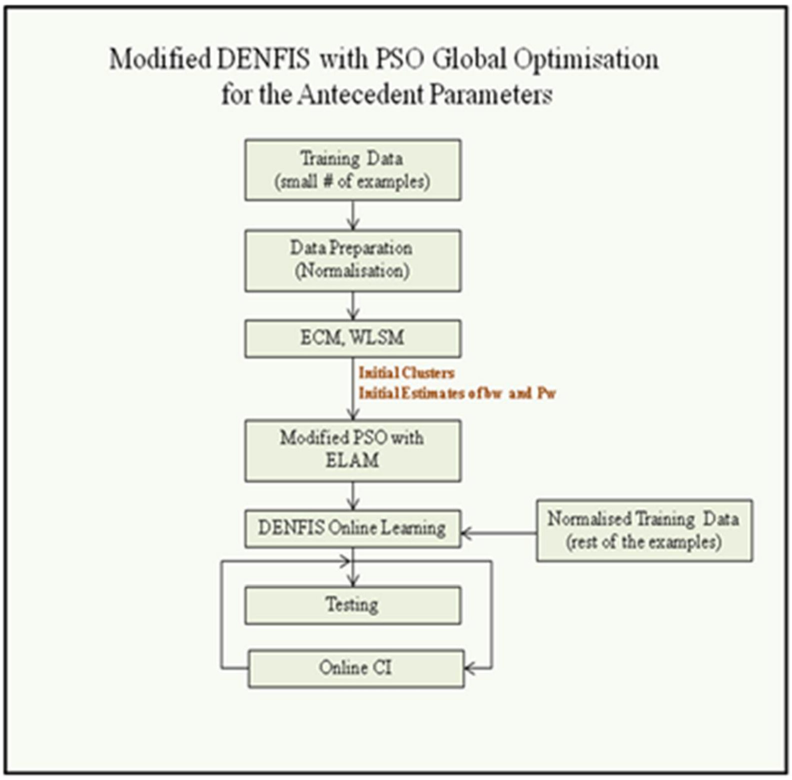


Figure 3-24 MDENFIS with PSO Global Optimisation

The same method was previously used to enhance the convergence rate in a neuro-fuzzy network application, in which the base particle was taken as the antecedent parameters of all created rules by the fuzzy entropy clustering method (FEC), the rest of the particles in the swarm were created by adding a small change to the base particle position to narrow the searching space around this region and enhance the learning rate (Lin et al. 2006). This approach has achieved better performance and learning speed than other genetic algorithm methods.

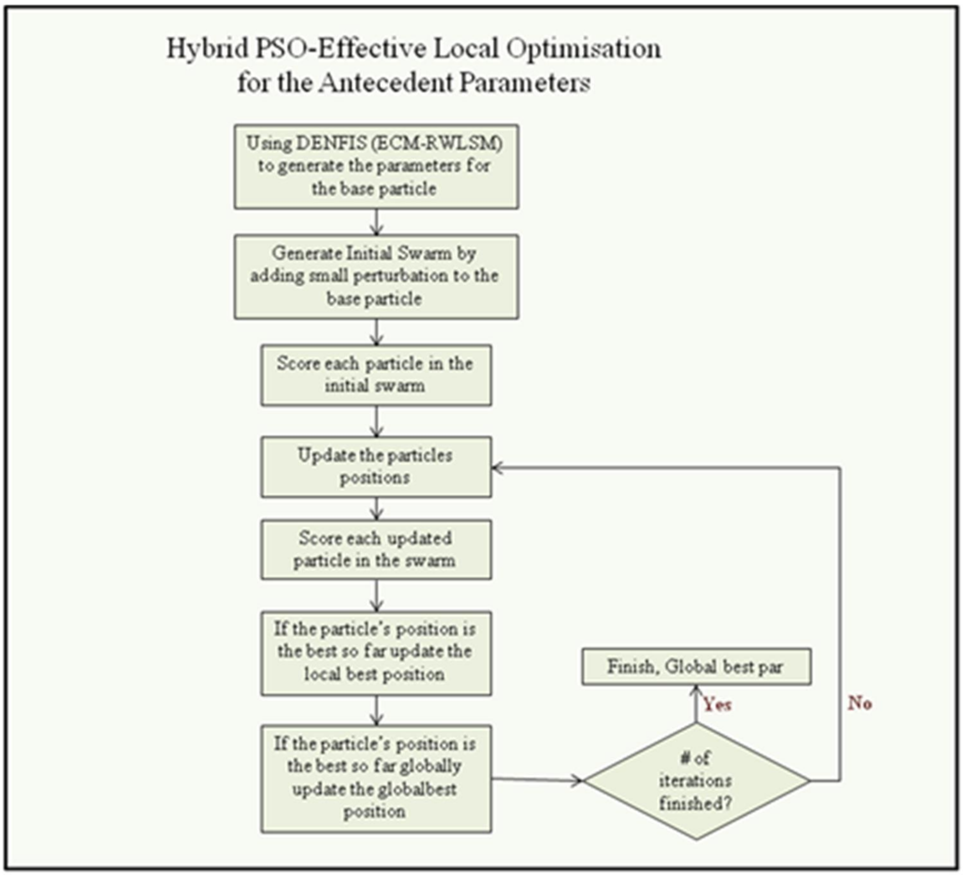


Figure 3-25 MPSO using Effective Local Approximation Method

As new data pairs become available new clusters might be added, existing clusters might be modified or existing clusters remain unchanged until the training stage finishes.

Three other changes to the DENFIS model are proposed:

1- While testing the model using incoming new input data, the model will also update itself automatically when the actual output becomes available using input-output data pair. The original model stops updating itself after the training stage. This means that new clusters could potentially be added not necessarily learned during the learning stage. The objective of this modification is to ensure that any transient genuine events are captured by the model.

2- To use all rules in the model instead of the m out of n generated rules proposed in the original DENFIS model. To take all the variations and degrees of membership of the data pair with all fired rules. Enhanced accuracy is obtain however on the cost of slightly longer processing time.

3- To change the used weight values in the original DENFIS model from weight value itself to 1-weight value. This will give more weight to the points closer to cluster centre.

The MDENFIS model explained above works online. As new inputs become available a single step ahead for the output is predicted, however if the future inputs are available this model will give multi-step ahead predictions as will be examined in the following two chapters 4 and 5. The tracking signal will still be used during this stage for the other health indicators to detect any early signs of drifting from the normal behaviour.

3.3.6 Long Term Fault Progression Prediction

The Remaining Useful Life also called prognostics can be found through a long term fault progression prediction model. The stopping or end point of the prediction horizon is normally a predetermined level like a trip level as shown in Figure 3-26.

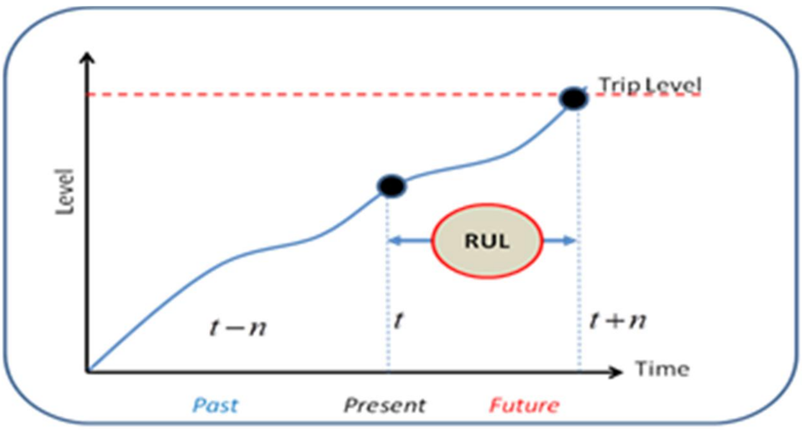


Figure 3-26 Remaining Useful Life (RUL) Prediction

Majority of the work in the literature in prognostics and long term forecasting assumed that the inputs are available, this is through splitting the full history dataset of the machine into two parts, one is used for training the model and one is used for testing it (forecasting future values). However in reality this isn't the case especially when you don't have a full history of progression from the early stages of development of the fault until failure, example being a newly installed machine. Add to that the complications of the uncertainty in the future operating regime of the machine, where variable demand is expected.

Taking the above into consideration, Figure 3-27 proposes two different prediction horizons for each of the inputs:

1- Short Term Prediction Using ARIMA
2- Long Term Prediction Using a hybrid of EMD and ARIMA

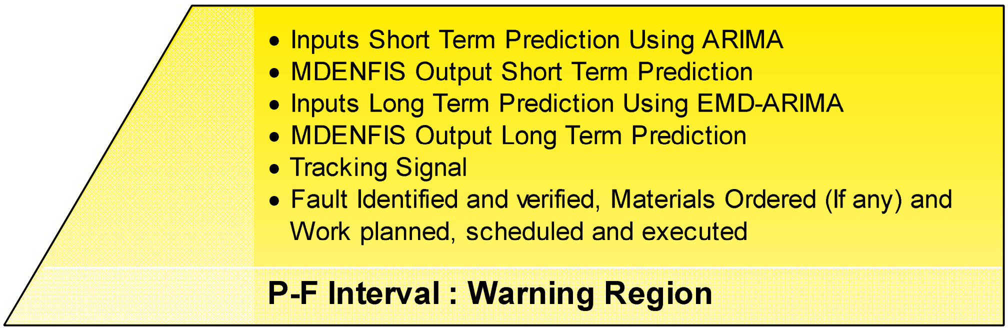


Figure 3-27 Warning Region Short Term and Long Term Prediction

The MDENFIS trained previously is used with the short term and long term predicted inputs to predict the output short term and long term, respectively. The tracking signal will still be used during this stage for the other health indicators to detect any early signs of drifting from the normal behaviour.

Having detected an abnormal behaviour of the machine outside its NOE, identified the source of this abnormal behaviour and predicted the long term fault progression. The next step would be to plan work and resources, order material if not in stock, schedule and execute the work as explained in Figure 4-28. Any major work done on the machine will require a new baseline to be completed such that this can be trended and monitored but in the first instance this is used to ensure that the work previously done is completed to a high standard and any concerns are addressed immediately.

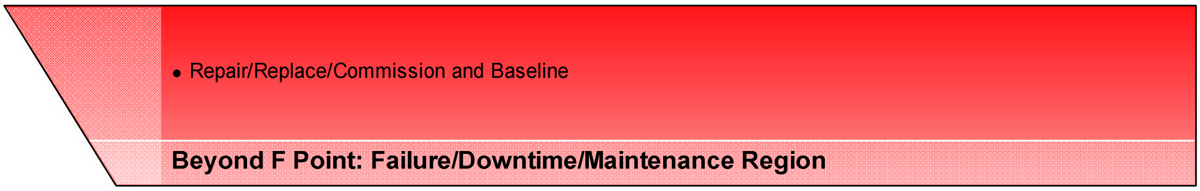


Figure 3-28 Failure/Downtime/Maintenance Region

3.4 SUMMARY

This chapter is first covered the enabling technologies for the integrated condition monitoring and prognostics proposed framework. This framework is fully aligned with ISO 17359:2011 (BS ISO 17359:2011 2011) and is built around the IP-PF curve. The novel integrated model is then presented in different stages through the life cycle of the machine, starting with the normal operating envelope stage (I-P interval) where the main concern is to ensure that the machine is running within its normal operating envelope and any drifting outside this envelope is detected early enough to take an action. The second stage covers the area where the machine is running outside its normal operating envelope and one or more faults are starting to develop (P-F interval). Within this interval the first thing to do is to identify the source of the abnormality in the behaviour through the fault classification and diagnostics step, predict the machine short and long term behaviour, and plan work and resources accordingly to enable timely intervention without compromising the integrity of the machine but in the same time not to intervene too early. The last step is to execute work, commission and baseline the machine and then the same process goes in cycle.

4 MODEL VALIDATION USING BENCHMARK DATASETS

4.1 Introduction

This chapter presents a number of case studies using benchmark datasets used by other researchers (Theophano 2010), (Kasabov and Song 2002), (Pratama et al. 2013), (Dovžan and Škrjanc 2011) and (Hwang and Song 2008) for comparison purposes and to proof capability of the proposed model. Before discussing the case studies, a short description of important steps when dealing with time series data is presented below.

Time series analysis (Data mining) includes different techniques to extract important information from the signals, Figure 4-1. This can be as simple as using data visualization methods such as XY plot however; real time signals are more complicated and as such extra work is required to better understand the internal structure of these signals. Statistical methods like calculating the mean, standard deviation, auto and partial correlation graphs are some of the tools used to understand the characteristics of the time series. Transforming the time domain data into a frequency domain using Fast Fourier Transformation can give vital information that might be hidden in the time domain data. More advanced techniques like principle component analysis (PCA), support vector machines (SVM) and clustering are used in pattern recognition for classification purposes and regression analysis.

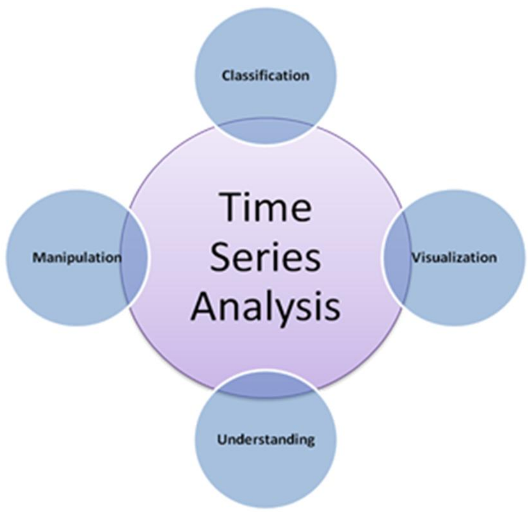


Figure 4-1 Time Series Analysis

Pre-processing the data prior to model implementation is an important step; the old paradigm "Garbage In-Garbage Out" is extremely dangerous in these situations. Outliers for example can shift the mean and standard deviation and can cause problems in modelling outcomes. Depending on the type of analysis sometimes removing outliers isn't the solution especially if these outliers will identify important things like the number of start-ups and shutdowns within certain period of time for example. Several statistical methods are used to detect outliers the most popular ones are the box plot rule (Tukey 1977) and Grubbs test (Grubbs 1950). Outliers can then be removed manually, by using a macro or using one of the smoothing techniques like the moving average and Fourier smoothing. Figure 4-2 shows an example for a temperature trend that was smoothed using the robust LOESS or LOWESS (locally weighted scatter plot smoothing) smoothing method with 0.1 span. This method fits smooth surfaces to the data presented using non parametric regression methods (Cleveland 1979). It uses locally weighted linear regression method to smooth the data. The robust option is used to make the manipulation process less sensitive to any outliers by introducing a weight function.

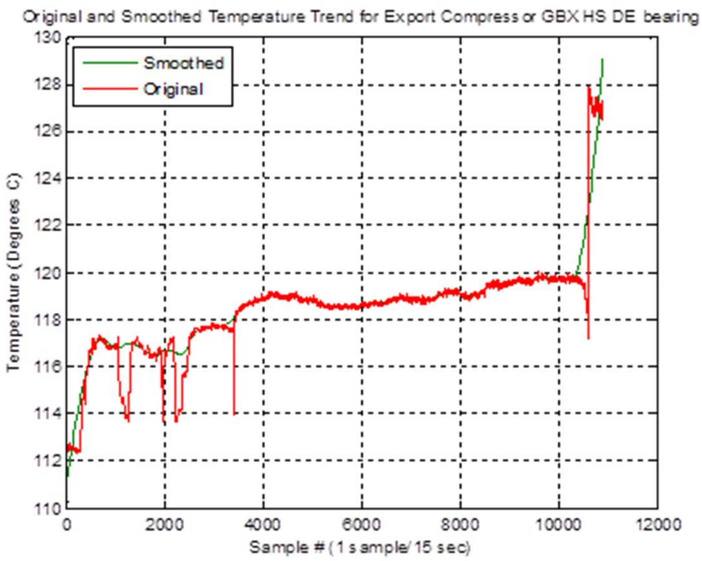


Figure 4-2 Original and Smoothed Temperature Trends

In addition to outliers, NaNs (Not a Number values) are very common as a result of communication and/or instrumentation problems and should be removed since some Softwares won't allow proceeding any further in the analysis without removing all NaNs from the data. The same goes to missing data.

Examples are shown below for outliers and NaNs discovered during the data cleaning stage, Figure 4-3:

	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
7403	17.3241863	23.997242	31.14444351	61.5971756	53.5458183	117.363731	65.76330566	62.85147095	60.88841[illegible]48	67.63240[illegible]4	62.17869186	59.11446762	53.57926941	54.40216064
7404	0.1999715	17.0270901	4.824598789	2.29386425	0.18741988	50.3165779	54.31359482	53.07974625	52.1140[illegible]19	DCS failed	52.91215515	51.51117706	38.65187073	39.34013367
7405	3.14121914	15.5274572	4.90999651	12.16001797	3.18319798	19.1028194	52.2040863	52.57523346	51.33703[illegible]5	DCS failed	52.11098862	50.80918503	28.3418026	29.03159904
7406	5.13260174	14.2777624	4.824598789	19.27788162	5.18528318	10.4878225	50.24296951	51.57510376	50.32683945	[illegible]	50.76963425	49.36233139	24.26874542	24.95501518
7407	17.915781	19.9951153	4.466710567	63.92052841	18.054781	8.49014187	48.40575409	50.34166336	49.12579727	DCS failed	49.33250427	47.81772995	25.20276833	25.89956474
7408	15.8727379	16.1210613	4.558817387	57.37799454	15.9931631	10.6126776	46.1588707	49.1223793	48.04512405	46.25703049	48.06257248	46.10047913	26.20680046	26.90019417
7409	17.5325012	24.6064892	29.51917458	63.0259819	50.6658401	113.992645	63.46883011	61.19651031	59.19276047	45.39240646	60.42661667	57.50911331	45.23159027	45.5454483
7410	15.9977198	24.4627743	30.00796127	58.15148544	51.090660[illegible]	117.[illegible]8412	63.84036255	61.09955978	59.08267975	40.89161682	60.43730164	57.4822464	51.14700699	52.03324127
7411	15.9810562	23.681715	30.3495369	57.08550644	49.47884[illegible]5	116.61[illegible]601	63.98116684	61.23250961	59.15940857	8.351659775	60.56404877	57.57429886	53.52680969	54.10997391
7412	16.6726246	24.2128353	30.32213211	59.51162338	51.2280[illegible]98	Error	64.16222382	61.49113083	59.4992485	45.27736664	60.80460358	57.89304733	54.17655945	54.76897812
7413	16.6901379	25.8999233	30.81708145	60.02662277	55.188[illegible]877	Error	66.16617584	63.15856552	61.25651169	42.03106689	62.4821701	59.58955002	54.54065323	55.61667252
7414	17.3908539	25.0813732	30.33309555	62.08104324	53.339[illegible]988	Error	65.91259766	62.94418335	60.92251205	[illegible]	62.29346848	59.32392883	54.31364822	54.99126053
7415	16.8359356	24.0191326	30.09647751	61.22289658	52.63[illegible]9956	Error	65.95256042	63.00881577	60.93157959	[illegible]	62.34220505	59.35183716	54.25190735	54.97475052
7416	16.6951256	23.9004116	30.32482719	60.33585739	52.04[illegible]2527	Error	66.00993347	63.05470657	60.95314407	-33.42773438	62.38540268	59.37791443	54.24524307	54.93716431
7417	16.8975925	24.2128353	30.48618507	60.04787064	52.7[illegible]7459	Error	66.02311707	63.0846138	61.04321289	28.91807747	62.3950386	59.41108704	54.24651718	54.97742844
7418	16.975914	24.7502041	30.82510757	60.85690689	53.17[illegible]2652	Error	66.03223419	63.07337952	61.05802536	35.90686417	62.41891861	59.44054031	54.44348907	55.25619507
7419	16.9800854	23.9004116	31.22838783	60.42677689	52.37[illegible]2387	Error	66.09184265	62.95269012	60.92074966	44.86820602	62.3443985	59.29133224	54.19945526	54.819664
7420	16.9475842	24.1503506	30.39059639	60.88233566	52.7[illegible]7459	Error	65.942276	62.83747101	60.76734543	46.61376953	62.23435974	59.18989563	54.05770111	54.81357574
7421	16.7984028	23.8598175	30.72097778	60.40443039	52.308[illegible]455	Error	65.48756409	62.44927216	60.40107727	55.16392899	61.80747986	58.77732468	53.48468399	54.37231064
7422	16.9042587	24.0191326	29.90801239	60.69535446	53.002[illegible]422	Error	65.23194885	62.11951828	60.0742836	49.85552216	61.52323151	58.47116089	53.26593781	54.42993927
7423	17.0259056	24.5315075	30.97581482	60.89779282	53.164[illegible]72	Error	65.22917175	62.08029938	60.0899086	53.76950836	61.52447128	58.5016861	53.44179916	54.59570313
7424	16.9059353	24.3065624	30.32901192	60.19172668	53.22740[illegible]4	Error	65.2522049	62.09119797	60.05647278	55.60652161	61.48435974	58.48921585	53.33031464	54.54886627
7425	17.0359058	24.406538	30.66014671	60.97304916	53.477138[illegible]	Error	65.02997589	61.83348465	59.80685043	50.56030273	61.21529388	58.18427658	53.03482056	54.45687866
7426	16.9809132	24.3721504	30.66022491	60.61712646	53.2959251	[illegible]	65.00009155	61.96121979	59.93503189	[illegible]450851	61.29071426	58.34170532	53.29445267	54.71622849
7427	16.8959255	24.0628719	31.36751366	60.06777191	52.3901024	116.864311	64.86219025	61.75896835	59.69607[illegible]25	51.624[illegible]461	61.11591339	58.10963058	53.08426666	54.34080887
7428	16.914257	24.525259	31.22706604	60.22548676	52.5400391	117.863152	64.87546539	61.89333725	59.860[illegible]816	41.13306[illegible]46	61.25611115	58.29084396	53.46541977	54.76155853
7429	10.2485399	19.4952374	4.649099827	37.02376175	10.3705664	63.913372	61.01752853	59.22326279	58.443[illegible]142	DCS failed	58.65474319	57.01655197	44.68931198	45.73983383
7430	10.2235432	11.8096151	4.649099827	37.06452179	10.3705664	35.8333931	56.05718231	56.48096848	54.99734[illegible]9	DCS failed	55.6679039	54.27456665	29.98660278	31.03986931
7431	10.1402216	10.8098593	4.649099827	36.76655197	10.3080931	25.5952816	53.42485809	55.23921967	53.58485031	DCS failed	53.86935043	52.58398438	24.81202316	25.86298561
7432	0.03332859	11.4034643	4.737661839	1.618917227	0.06247329	12.8600683	51.47732544	53.96224594	52.27873993	DCS failed	52.26831436	51.09111404	25.33611488	26.42666435

Figure 4-3 Example Database Showing Outliers and NaNs

In steady state applications all transient events (during start-up and shutdown) aren't wanted and ideally filtering this data is based on certain operating tags that identify a steady state running condition like the motor current (A) and speed (RPM).

The second step in data pre-processing is data normalization where the data is normally scaled to fall within a specific range like [0, 1]; Two methods are available for normalization the min-max, equation 4-1 and the z-score normalization methods, equation 4-2 (Theophano 2010).

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \text{is the min-max normalisation method} \tag{4-1}$$

$$x_{norm} = \frac{x - \mu}{\sigma} \quad \text{is the z-score normalisation method} \tag{4-2}$$

where
x is the actual value
$x_{\min}$ is the minimum value in the range
$x_{\max}$ is the maximum value in the range
μ is the mean value
σ is the standard deviation

The min-max normalization method is used to normalize the temperature trend in Figure 4-4.

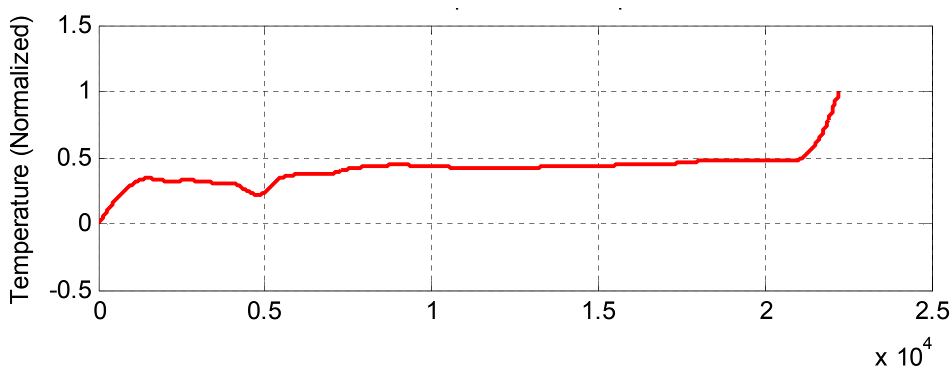


Figure 4-4 Temperature Trend Normalisation using Min-Max Method

Normalizing the data is very important when comparing different variables with different units as this will scale all variables to the same range making it easier to find any similarity in the behaviour. Another application is in Neural Networks and Fuzzy Logic where the input data is required to lie within a [0, 1] range as this will enhance the forecasting accuracy.

4.2 TESTING PARAMETERS

4.2.1 Testing Datasets

The following three datasets are used in the literature as benchmark datasets in the areas of neural networks, fuzzy logic and hybrid system integrating both neural networks and fuzzy logic and are specifically used for time series prediction and classification applications. To enable performance comparing the proposed model with other researchers' work, the same datasets will be used here to test the implementation:

1. **Mackey-Glass Time Series Dataset:**

   A chaotic time series developed to model the white blood cells reproduction (Mackey and Glass 1977). Represented by the following differential equation 4-3 and can be solved using the 4th order Runge-Kutta numerical method in Matlab, Figure 4-5.

$$\frac{dx(t)}{dt} = \frac{0.2x(t-tau)}{(1+x(t-tau)^{10})} - 0.1x(t) \qquad \textbf{(4-3)}$$

Where, $x(t)$ is the density of mature cells in blood circulation, tau is the delay constant with a given value of 17 and $x(0)$ is the initial condition equal to 1.2.

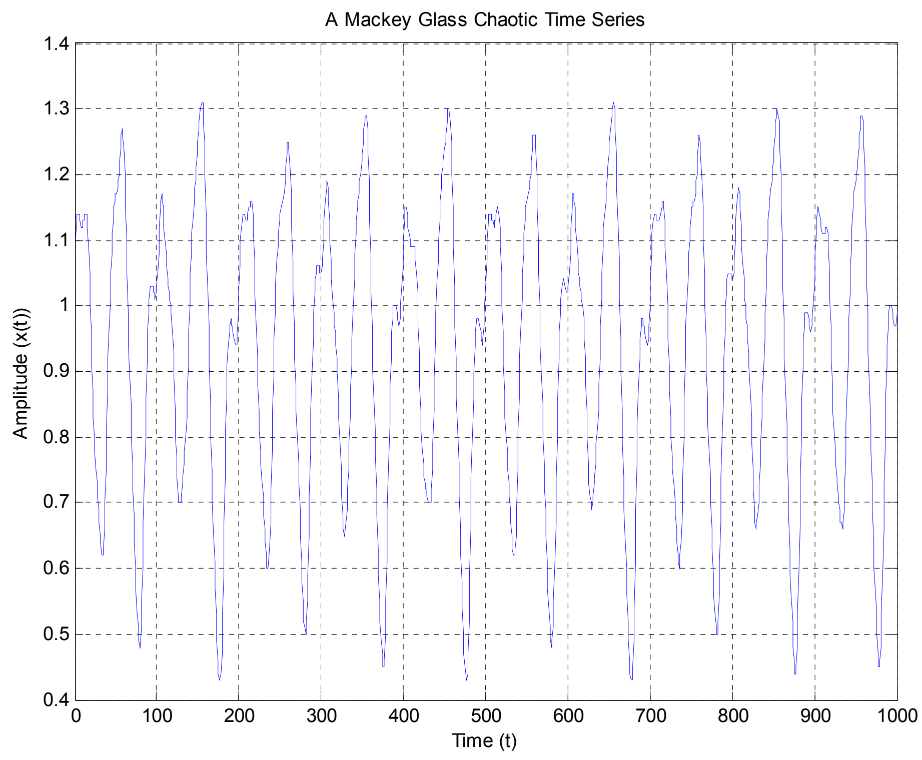


Figure 4-5 Mackey-Glass Time Series

2. **Gas Furnace Dataset: Box & Jenkins Series J**

   The gas furnace dataset (Box and Jenkins 1976) and (Box et al. 1994) contains two inputs and one output. A total of 292 inputs/output pairs. The process itself is a burning process for methane in a furnace and the CO_2

concentration is the output from the furnace. The first input used is the methane flow rate at t-4 (where t is the present time step) and the second input is the CO_2 concentration at the present time (t) and the output is the CO_2 concentration at t+1 i.e. one step ahead, Figure 4-6.

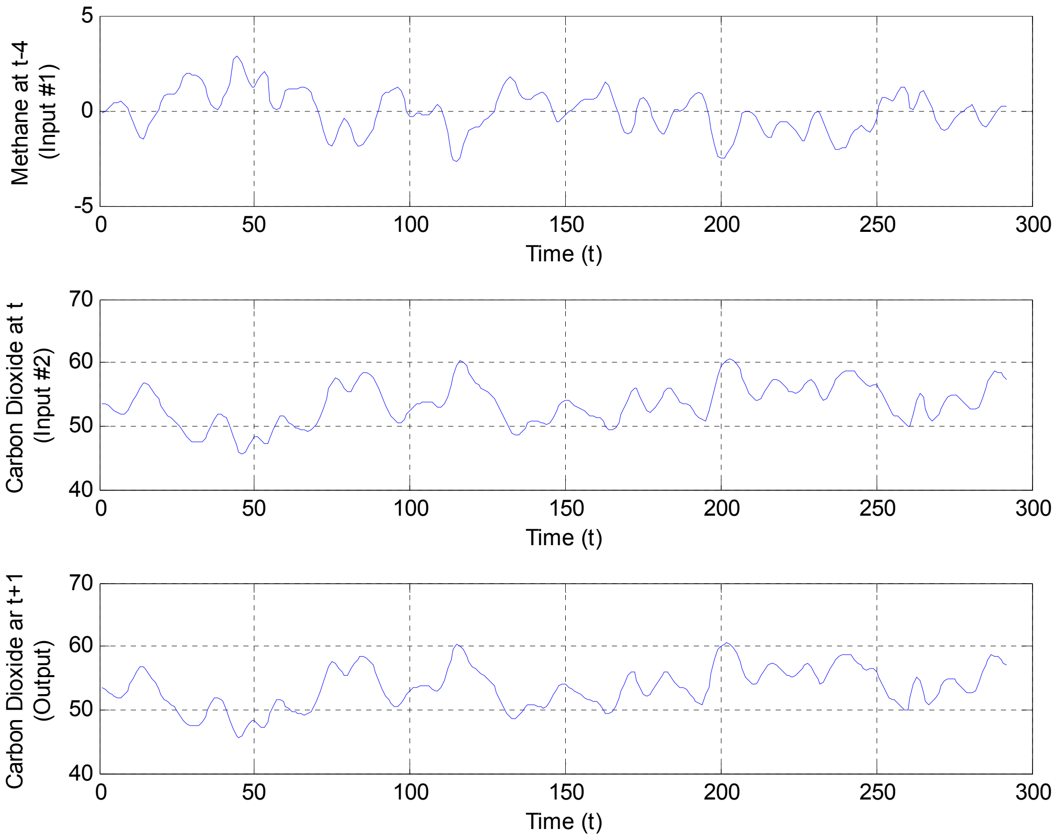


Figure 4-6 Gas Furnace Dataset Inputs/Output

3. **<u>Iris Flower Dataset:</u>**

This dataset (Fischer 1936) is used for classification problems. It contains 50 samples of each of the Iris Flower species, namely: Setosa, Virginica and Versicolor. The Sepal and Petal length and width are used as inputs and the flower species is used as the output. Figure 4-7 shows the 4 inputs characteristics and 1 output (Iris Species). Figure 4-8 shows the same data but as three different clusters in green, blue and red referring to Iris Setosa, Virginica and Versicolor respectively. This data is clustered using k-means clustering method.

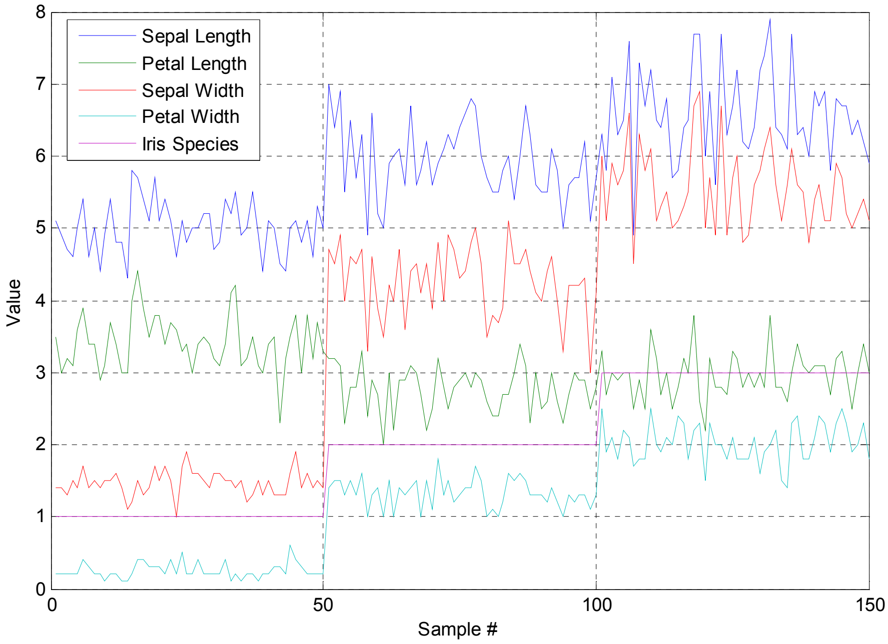


Figure 4-7 Iris Dataset Inputs/Output Trends

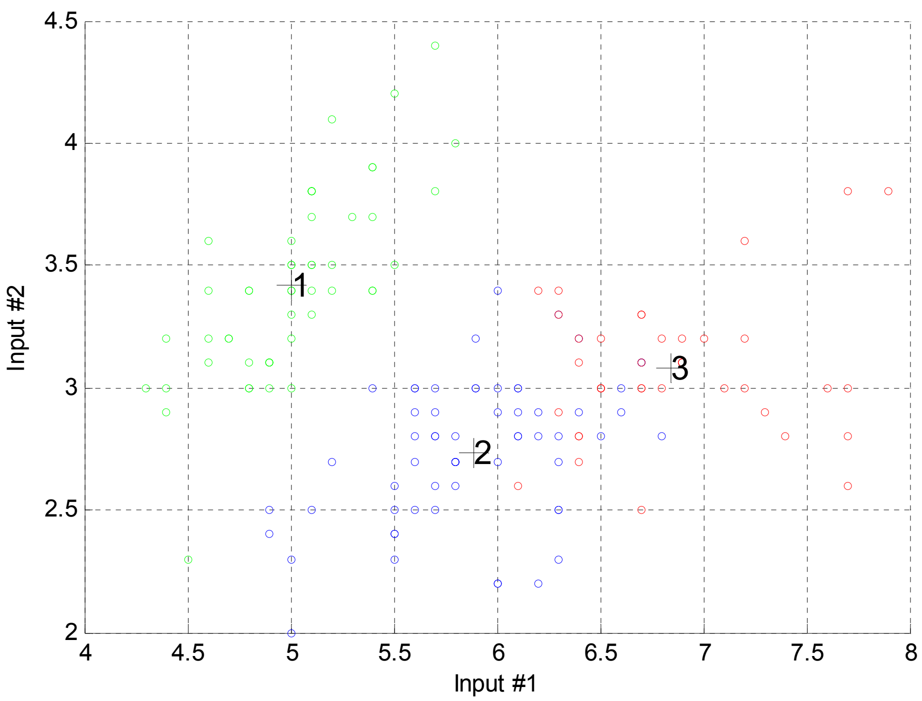


Figure 4-8 Iris Setosa, Virginica and Versicolor Clusters (1-3)

4.2.2 Performance Metrics

Performance Metrics are used to measure the model performance in the main areas of interest, like the processing speed using different number of inputs, size of dataset and complexity of the input signals, prediction accuracy and the ability to resist sudden changes in certain inputs due to instrumentation problems for example. The following metrics were previously used by other researchers to measure their models performance and will be used in this research for comparison purposes, references will be made in each of the case studies presented for their usage accordingly:

1- The **Mean Square Error (MSE)**: Described by equation 4-4.

$$MSE = \frac{1}{n}\sum_{l=1}^{n}(actual - predicted)^2 \quad \textbf{(4-4)}$$

2- The **Root Mean Square Error (RMSE)**: Described by equation 4-5.

$$RMSE = \sqrt{MSE} \quad \textbf{(4-5)}$$

3- The **Mean Absolute Error (MAE)**: Described by equation 4-6.

$$MAE = \frac{1}{n}\sum_{l=1}^{n}\left|actual - predicted\right| \quad \textbf{(4-6)}$$

4- **Robustness**: The ability of a model to resist changes in one of the parameters whilst the rest are stable. i.e. Step change in one parameter due to instrumentation error. The following equation is used to measure Robustness with a 0 being the target. i.e. the model will not follow (over fit) the change, equation 4-7:

$$\text{Robustness} = \frac{mean\ (x_0 - x_\Delta)}{\Delta} \quad \textbf{(4-7)}$$

Where x_0 is the model ideal estimate of the parameter, x_Δ is the model estimate when delta change is applied and Δ is the delta change applied to the parameter

5- **Spillover**: The ability of a model to resist changes in the model's parameters as a result of changes to one parameter, equation 4-8.

$$\text{Spillover}_j = \frac{1}{\text{M - 1}} \sum_{i=1, i \neq j}^{M} \frac{mean(x_{i0} - x_{i\Delta j})}{\Delta_j} \quad \textbf{(4-8)}$$

Where x_{i0} is the model ideal estimate of the parameter when no parameter changes is applied , $x_{i\Delta j}$ is the model estimate of parameter i when delta change is applied to parameter j and Δ_j is the delta change applied to the parameter j. M is the total number of parameters.

6- **Error**: The error measure used here is the root mean squared error divided by the standard deviation of the actual data, equation 4-9.

$$\text{Error} = \frac{rms(error)}{\sigma_x} \quad \textbf{(4-9)}$$

σ_x is the standard deviation of the actual data x.

7- **Non-Dimensional Error Index (NDEI):**

$$\text{NDEI}=\frac{RMSE}{\sigma_x} \quad \textbf{(4-10)}$$

σ_x is the standard deviation of the actual data x .

8- **Processing Time:** Processing time is an important measure especially when dealing with online models where data feed is coming continuously at a certain frequency. The algorithm efficiency is more meaningful when tested on a large number of machines rather than a single case, however as the benchmark datasets were utilised for testing the processing time of various algorithms, the same are used here for comparison purposes and for testing the proposed algorithm’s efficiency.

4.3 CASE STUDY No.1

4.3.1 Hypothesis

Prediction Accuracy Enhancement: Reviewing the literature, it is claimed that DENFIS outperformed the performance of some of the other well-known models including for example: Multiple Linear Regression (MLR), Radial Basis Function (RBF) and Evolving Fuzzy Neural Networks (EFuNN). Since this model is the basis for the MDENFIS proposed in this research it is important to proof this hypothesis first.

4.3.2 Testing

For the purpose of testing the above hypothesis the Gas Furnace dataset benchmark dataset is used. The dataset is splitted into two parts: one for training the models (80% of the dataset, 235 input/output pairs) and the other for testing the models (20% of the dataset, 58 input/output pairs). Two models are selected for comparison with the DENFIS model, the multiple Linear Regression and Radial Basis Function models. To compare the results obtained from all three runs the Mean Square Error, Root Mean Square Error and Mean Absolute Error are compared for both the training and testing predictions. Results are included in section 4.3.3.

4.3.3 Results and Discussion

The Carbon Dioxide Concentration one step ahead output is plotted for comparison purposes with the training and testing predictions of the MLR, RBF and DENFIS models in Figure 4-9. The performance metrics in the form of MSE, RMSE and MAE are shown in Figure 4-10 and Table 4-1. When comparing the training prediction accuracy measures, better accuracy is achieved with the DENFIS model with at least 10% improvement. The testing prediction accuracy results aren't comparable, with the best accuracy achieved with DENFIS and the worst with RBF (DENFIS RMSE error is twice as good as the RBF RMSE error and 2% better than MLR).

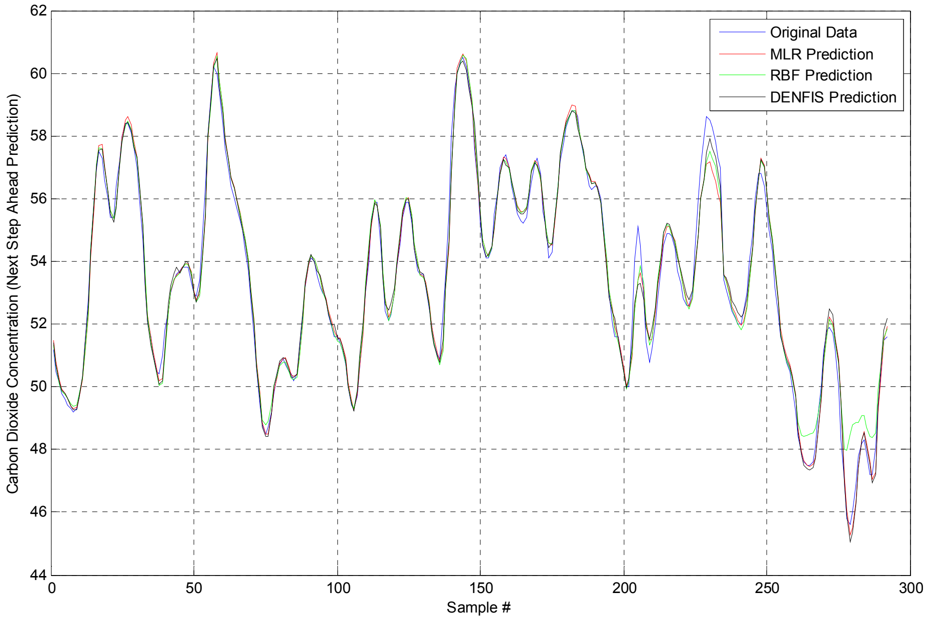


Figure 4-9 CO2 Concentration Prediction One Step Ahead

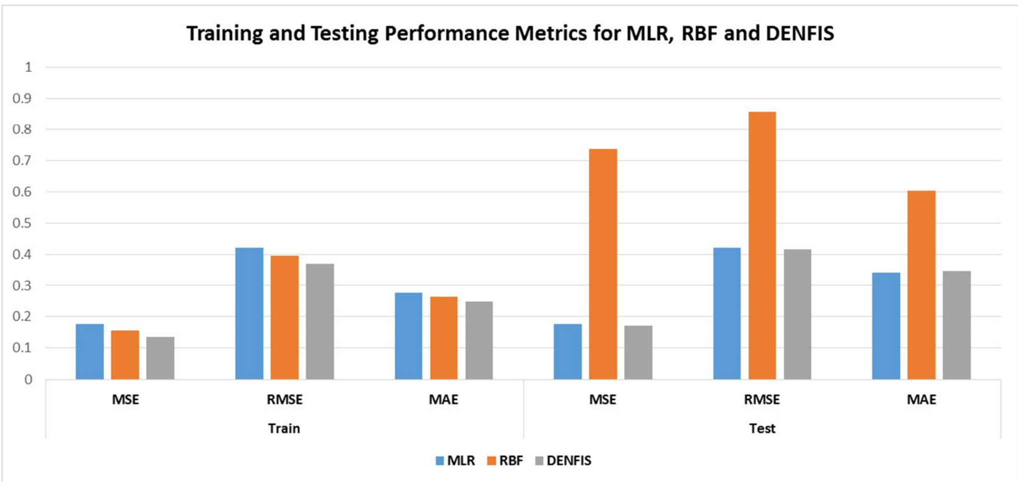


Figure 4-10 Training and Testing Performance Metrics for MLR, RBF and DENFIS

Table 4-1 Training and Testing Performance Metrics for MLR, RBF and DENFIS

Model	Train			Test		
	MSE	RMSE	MAE	MSE	RMSE	MAE
MLR	0.1767	0.4204	0.2775	0.1776	0.4215	0.3413
RBF	0.1559	0.3948	0.2648	0.7354	0.8575	0.6028
DENFIS	0.1359	0.3686	0.2497	0.1719	0.4146	0.3462

The Gas furnace dataset 2 is used to proof the first hypothesis. However, DENFIS was also compared to other methods in the literature using Mackey-Glass Chaotic Time Series dataset 1. For example, DENFIS online and offline models were compared with the resource-allocating network (RAN), evolving fuzzy-neural network (EFuNN), evolving self-organising maps (ESOM), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear perceptron (MLP) (Kasabov and Song 2002) and better testing NDEI and training time in seconds were found, Table 4-2. The first 4 rows show tests done online and the remaining tests were done offline. DENFIS was at least 10% better than the other methods when comparing testing NDEI for the online tests and 50% better than the other methods when comparing the testing NDEI for the offline tests add to that the shorter processing time.

DENFIS was also compared to evolving Takagi Sugeno (eTS), Simple eTS and Bayesian ART-based Fuzzy Inference System (BARTFIS) as shown in Table 4-3 and better testing NDEI was obtained with slightly longer processing time however comparable when for instance compared to Fast and Accurate Online self-organizing scheme for Parsimonious Fuzzy Neural Networks (FAOS-PFNN), Dynamic Fuzzy Neural Network (DFNN) and Generalised Dynamic Fuzzy Neural Network (GDFNN) (Pratama et al. 2013).

Table 4-2 Mackey-Glass Dataset Testing NDEI and Training Time

#	Method	Number of Fuzzy Rules	Testing NDEI	Training Time (s)
1	RAN	113	0.373	na
2	ESOM	114	0.32	na
3	EFuNN	193	0.401	na
4	DENFIS I	58	0.276	na
5	MLP-BP	60	0.090	1779
6	ANFIS	81	0.033	23578
7	DENFIS II	58	0.016	351

Table 4-3 Testing NDEI and Training Time Comparison

#	Method	Number of Fuzzy Rules	Testing NDEI	Training Time (s)
1	DENFIS	58	0.278	4.1
2	eTS	99	0.356	3.8
3	BARTFIS	24	0.301	3.66
4	FAOS-PFNN	44	0.1685	129.89
5	DFNN	18	0.1345	200.56
6	GDFNN	11	0.0959	190.67

Another comparison was made in (Dovžan and Škrjanc 2011), where 4 inputs were used including the current, 6 steps, 12 steps and 18 steps back to predict 85 steps ahead. DENFIS outperformed the prediction accuracy of the following models: eTS, RAN, ESOM, EFuNN, Neural gas and 3 of the rec.FCM models when comparable number of clusters are used, Table 4-4.

Table 4-4 Number of Fuzzy Rules and Testing NDEI Comparison

#	Method	Number of Fuzzy Rules	Testing NDEI
1	DENFIS	58	0.276
2	eTS	113	0.0954
3	RAN	113	0.375
4	ESOM	114	0.32
5	EFuNN	193	0.401
6	Neural Gas	1000	0.062
7	Rec.FCM	58	0.3085

4.4 CASE STUDY No.2

4.4.1 Hypothesis

Prediction Accuracy Enhancement: It is expected that the model prediction accuracy will be enhanced by using all the generated rules instead of the m out of n rules proposed in the original DENFIS model. It is also expected that further enhancement will be obtained by using exponential membership functions instead of the rectangular membership functions used in the original DENFIS model.

4.4.2 Testing

For the purpose of testing the above hypothesis Mackey-Glass Time Series dataset 1 is used. The objective is to predict the future value $x(t+85)$ from the following four inputs: $x(t-18)$, $x(t-12)$, $x(t-6)$ and $x(t)$. The first 200 points of the time series were disregarded. Starting from data point 201, 3000 consecutive data points were used for learning the model and 500 data points for testing the model (starting from $t = 5001 \text{to} 5500$). The same test was performed in (Kasabov and Song 2002) and results from this test are compared to results of testing the above hypothesis.

4.4.3 Results and Discussion

The following parameters within the DENFIS model requires optimisation to select the right parameters:

1. Initial number of learning data pairs from the learning dataset
2. Number of data pairs close to a cluster centre used to calculate the least square solution. Usually this is more than the number of input variables.
3. Sigma: The width of the exponential membership function.

Figure 4-11, shows the testing dataset including 4 inputs and one output. Figures 4-12, 4-13 and 4-14 show the training NDEI results for optimising the values of the initial number of learning data pairs, number of data pairs close to a cluster centre and sigma, in order. The number of rules used with this run for the MDENFIS is 30 rules, testing NDEI is 0.1925 and 2.028 s total processing time. Table 4-5 shows a comparison between this run and some of the other known models used in (Kasabov and Song 2002) and (Pratama et al. 2013). MDENFIS outperformed the performance of DENFIS, eTS, EFuNN and BARTFIS by at least 8% in terms of the testing NDEI and 40% quicker than BARTFIS in terms of the processing time even though the number of rules used was higher than BARTFIS (30 compared to 24 rules).

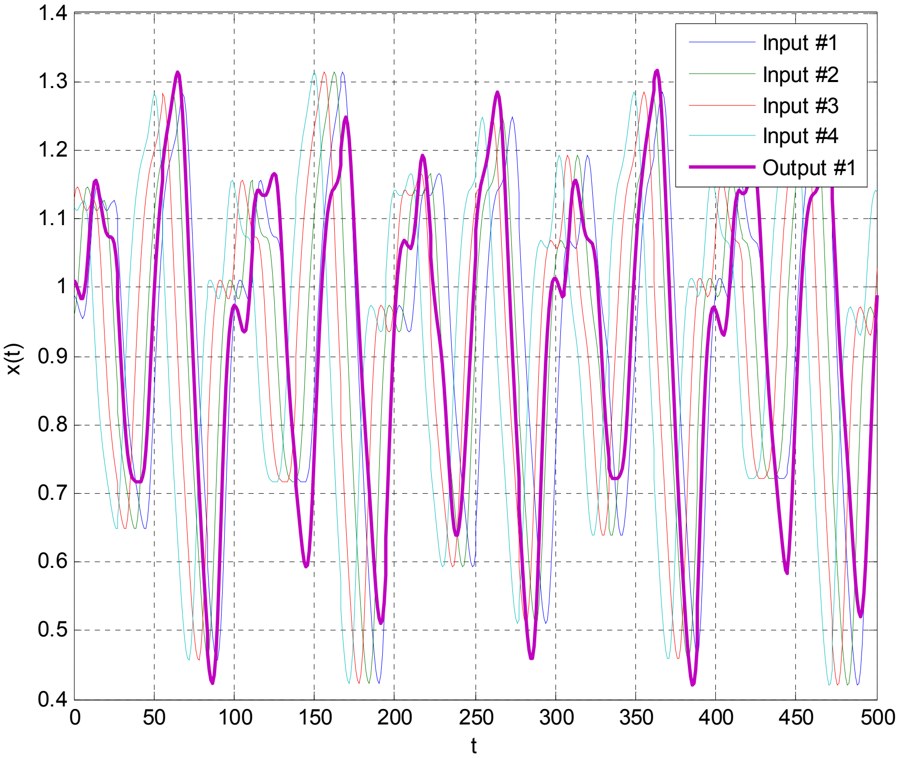


Figure 4-11 Mackey-Glass Dataset 1 Inputs/Output Trends

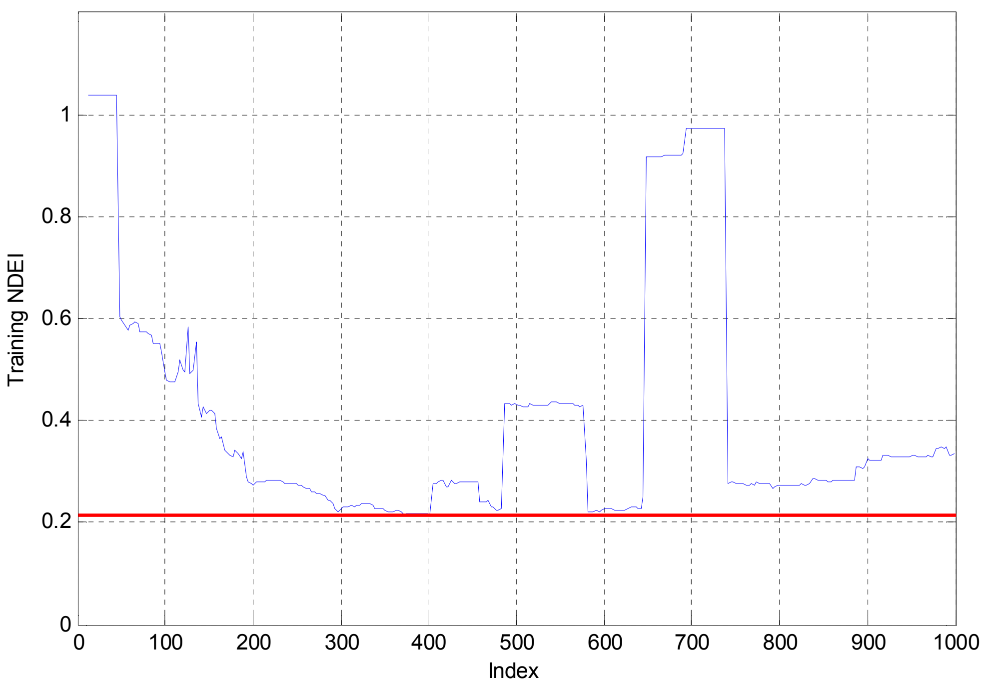


Figure 4-12 Initial Number of Learning Inputs/Output Pairs Optimisation

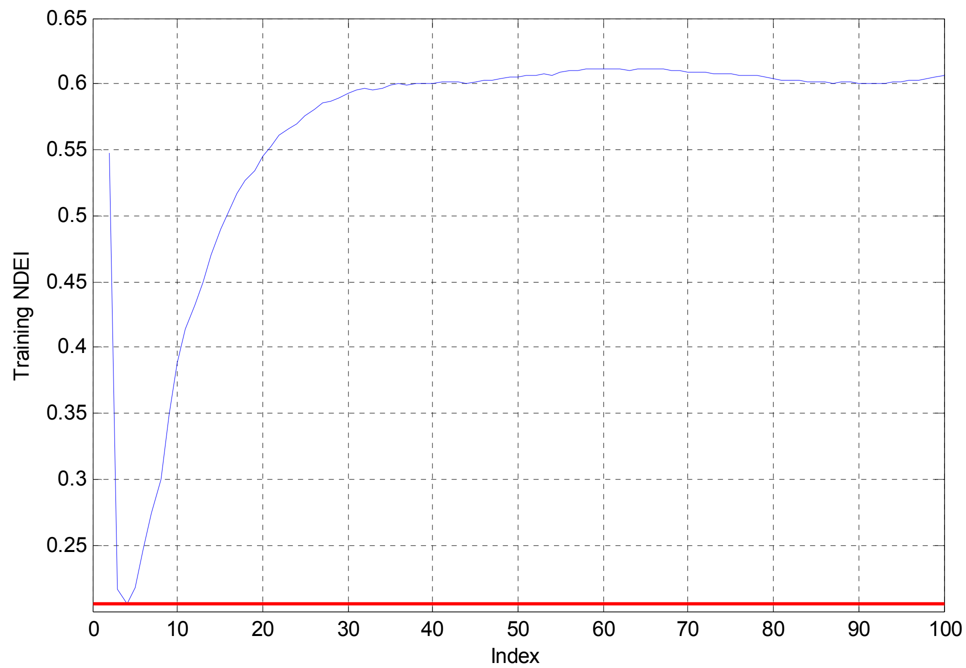


Figure 4-13 Initial Number of data pairs Close to a Cluster Centre Optimisation

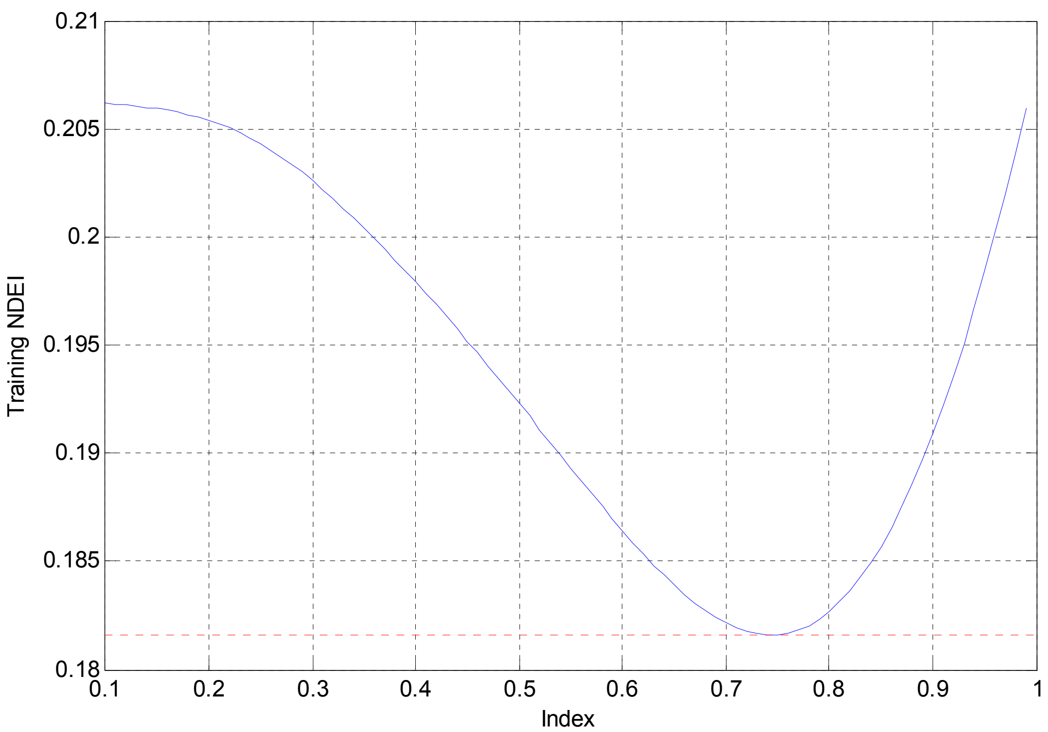


Figure 4-14 Initial Sigma Optimisation

Table 4-5 Comparison of Number of Rules, Testing NDEI and Training Time

#	Method	Number of Fuzzy Rules	Testing NDEI	Training Time (s)
1	MDENFIS	30	0.1925	2.028
2	DENFIS	58	0.278	4.1
3	eTS	99	0.356	3.8
4	EFuNN	193	0.401	na
5	BARTFIS	24	0.301	3.66

4.5 CASE STUDY No.3

4.5.1 Hypothesis

Prediction Accuracy Enhancement and Processing Time: It is expected that the model prediction accuracy will be enhanced by using an optimisation method like Particle Swarm Optimisation (PSO) to select the best cluster centres and width of the exponential membership functions. Using the initial cluster centres obtained with ECM as the first particle position in the PSO algorithm instead of some randomly selected centres and width will speed up the convergence time of the PSO.

4.5.2 Testing

For the purpose of testing the above hypothesis Mackey-Glass Time Series dataset 1 is used. The objective is to predict the future value $x(t+6)$ from the following four inputs: $x(t-18)$, $x(t-12)$, $x(t-6)$ and $x(t)$. The first 117 points of the time series were disregarded. Starting from data point 118, 500 consecutive data points were used for learning the model and 500 data points for testing the model. The same test was performed in (Hwang and Song 2008) and results from this test are compared to results of testing the above hypothesis. The following runs are conducted and the number of rules, training NDEI and testing NDEI are compared with the results in (IEC 60812 2006), the test results are shown in Table 4-6:

1- MDENFIS (I): MDENFIS with constant sigma (width of the exponential membership functions)
2- MDENFIS (II): MDENFIS with constrained ECM and constant sigma (same as above value)
3- MDENFIS (III): MDENFIS with constrained ECM and variable sigma (taken as the modified clusters radii following the constrained optimisation)
4- MDENFIS (IV): Same as MDENFIS (I) but with lower Dthr.
5- MDENFIS (V): Same as 4 but with MPSO. 10 Particle and 250 generations are used in the MPSO.

4.5.3 Results and Discussion

The first method MLP-BP showed extremely bad performance (at least 60% worse than the next better method in the list) when compared with the other 12 methods in Table 4-6. Taking this method out of the comparison, Figure 4-15 shows the remaining 11 methods used for the comparison. MDENFIS (I) with constant sigma showed better performance when compared to MLP-BP method 2 and ANFIS methods 3 & 4. However, due to lower number of rules, it is expected that this method has the shortest processing time. MDENFIS (IV) is the same as MDENFIS (I) however the Dthr value is smaller in order to obtain same number of rules as in methods 6 and 7. When compared with method 1-7 MDENFIS (IV) outperformed all of them by comparing both training and testing NDEI. MDENFIS (II) and (III) show similar performance. The last method 12, MDENFIS (V) using comparable number of rules to DENFIS I & II and DyNFIS showed better performance than all methods and an improvement of over 60% for the training NDEI and over 50% for the testing NDEI when compared with DENFIS (I) having in mind the number of rules for DENFIS I is 883 compared to 55 rules for MDENFIS (V). Figure 4-16 shows the training NDEI trend over 250 generations of the MPSO and Figure 4-17 shows the original cluster centres obtained using the ECM and the modified cluster centres using MPSO plotted in the same plot as x(t-18) and x(t-12). Figure 4-18 shows the original and predicted test output in addition to the error signal (difference between original and predicted output signals).

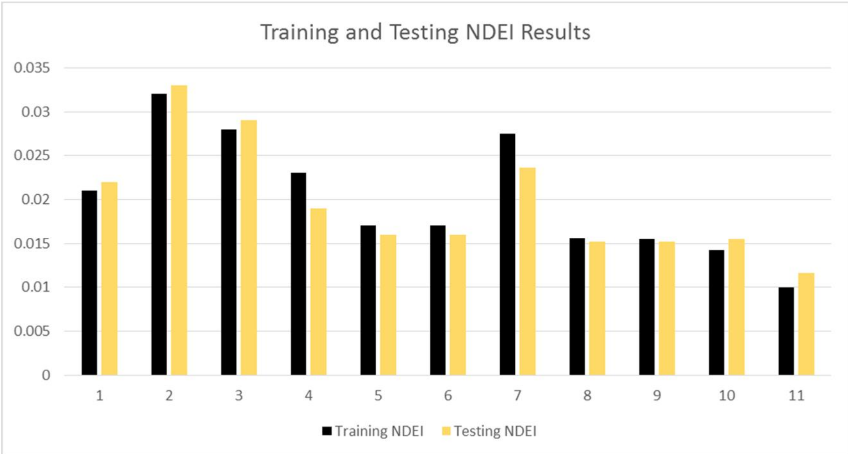


Figure 4-15 Training and Testing NDEI Results

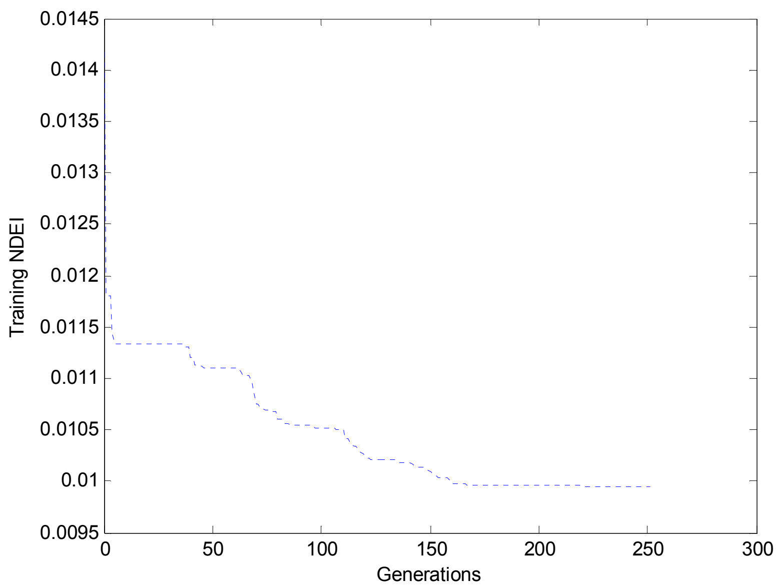


Figure 4-16 Training NDEI Trend Over 250 Generations of the MPSO

Table 4-6 Number of Rules, Training and Testing NDEI Testing

#	Method	Number of Rules	Training NDEI	Testing NDEI
1	MLP-BP	60	0.083	0.090
2	MLP-BP	60	0.021	0.022
3	ANFIS	81	0.032	0.033
4	ANFIS	81	0.028	0.029
5	DENFIS I (TS)	883	0.023	0.019
6	DENFIS II (MLP)	58	0.017	0.016
7	DyNFIS (TS)	55	0.017	0.016
8	MDENFIS (I)	35	0.0275	0.0236
9	MDENFIS (II)	35	0.0156	0.0152
10	MDENFIS (III)	35	0.0155	0.0152
11	MDENFIS (IV)	55	0.0142	0.0155
12	MDENFIS (V)	55	0.0100	0.0116

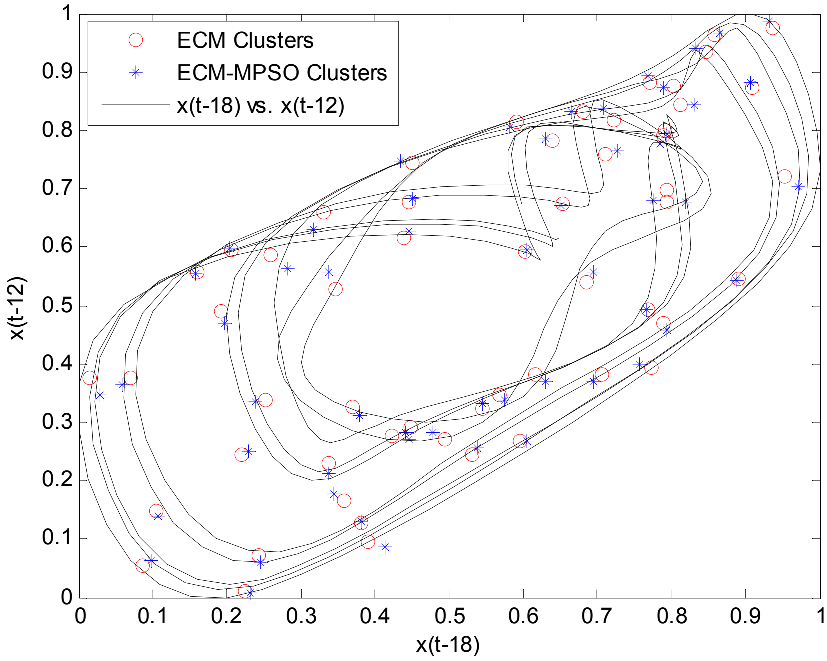


Figure 4-17 Original and Modified Cluster Centres

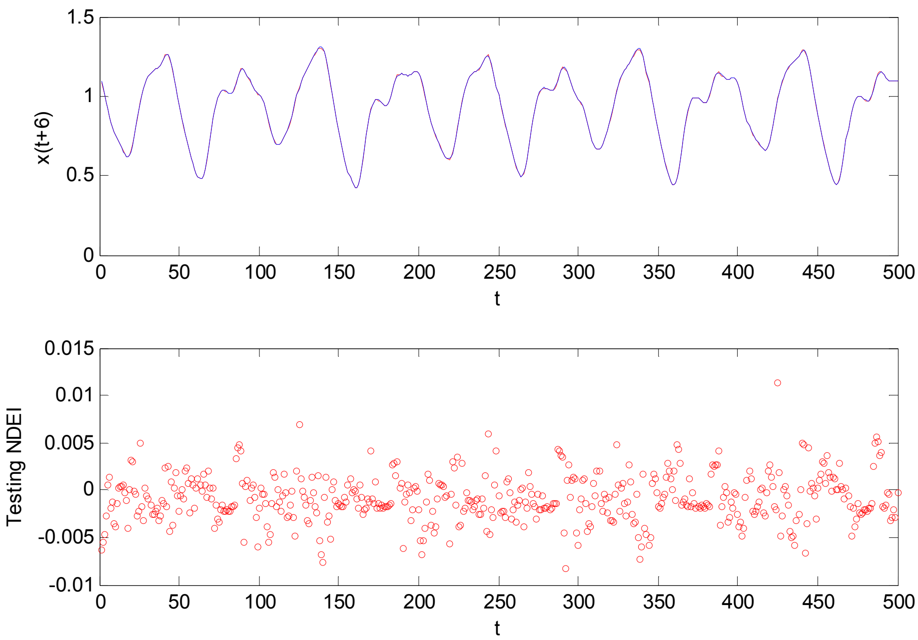


Figure 4-18 Original and Predicted Output at x(t+6) and the Testing Error

4.6 CASE STUDY No.4

4.6.1 Hypothesis

Ability to detect any sudden changes: It is expected that the model can detect any sudden changes in the trend. The model has capability to learn/add new rules while working online during the testing stage. However this is a constrained process depending on the objective of the model itself.

4.6.2 Testing

The gas furnace dataset 2 is used for this test. Total of 292 data pairs are used with two inputs, the first input is the methane flow rate at t-4 and the second input is the CO_2 concentration at the present time (t). The objective is to predict the CO_2 concentration at t+1 i.e. one step ahead prediction. The data is splitted into training dataset (the first 200 data pairs) and testing dataset (92 data pairs). The model is first tested using the original DENFIS model with minor changes (using exponential membership functions, different weight measure and using all fired fuzzy rules in the calculations), then the model is tested using the MDENFIS models.

4.6.3 Results and Discussion

Table 4-7 shows the MSE, RMSE, NDEI and number of rules for the training, testing and online update- testing stages. 100 data pairs are selected during the initial training stage. MDENFIS shows better performance at this stage with same number of rules 9. During the second stage of training i.e. online training, two additional rules are added (total of 11 rules), again MDENFIS outperformed DENFIS during this stage. The remaining 92 data pairs are then used to test the trained models (DENFIS and MDENFIS without the online update capability). Same number of rules 11. Better performance is achieved using MDENFIS during this stage. Finally, MDENFIS with online update capability is used to test the trained MDENFIS model. Additional 3 rules are added. The MSE, RMSE and NDEI obtained are better than using MDENFIS without the online update capability.

Table 4-7 MSE, RMSE, NDEI and Number of rules for the training, Testing and Online Update- Testing Stages.

	Training stage 1 (100 data pairs)		Training stage 2 (100 data pairs)		Testing		Online Testing (MDENFIS)
	DENFIS	MDENFIS	DENFIS	MDENFIS	DENFIS	MDENFIS	
MSE	0.0007	0.00052	0.0008	0.00046	0.0043	0.0033	0.0031
RMSE	0.0267	0.0230	0.0291	0.0215	0.0653	0.0572	0.0558
NDEI	0.1172	0.0999	0.1581	0.1172	0.2757	0.2411	0.2356
Rules#	9	9	11	11	11	11	14

Figure 4-19 shows the results of the clustering optimisation, the arrows in the figure are pointing towards the new locations of the clusters and Figure 4-20 shows the testing results.

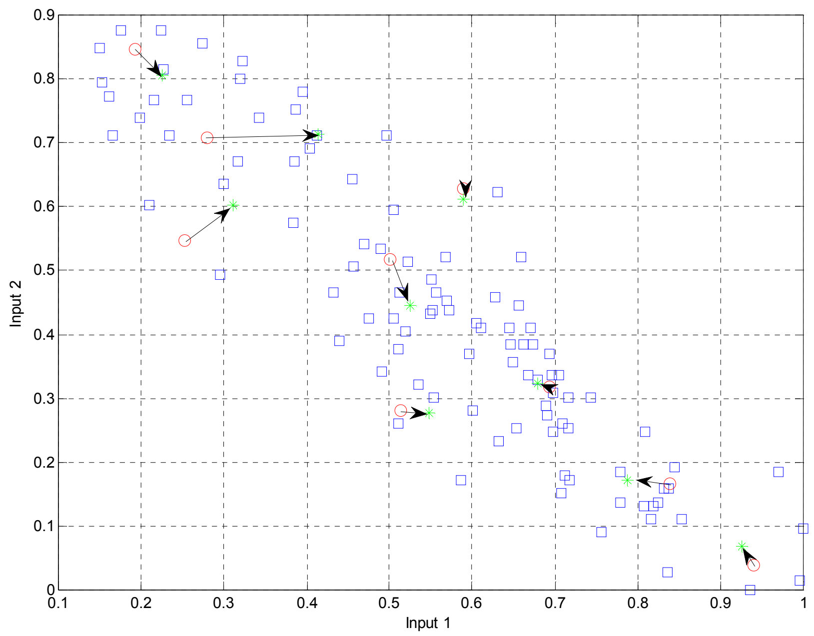


Figure 4-19: Gas Furnace Input Data Clustering Optimization

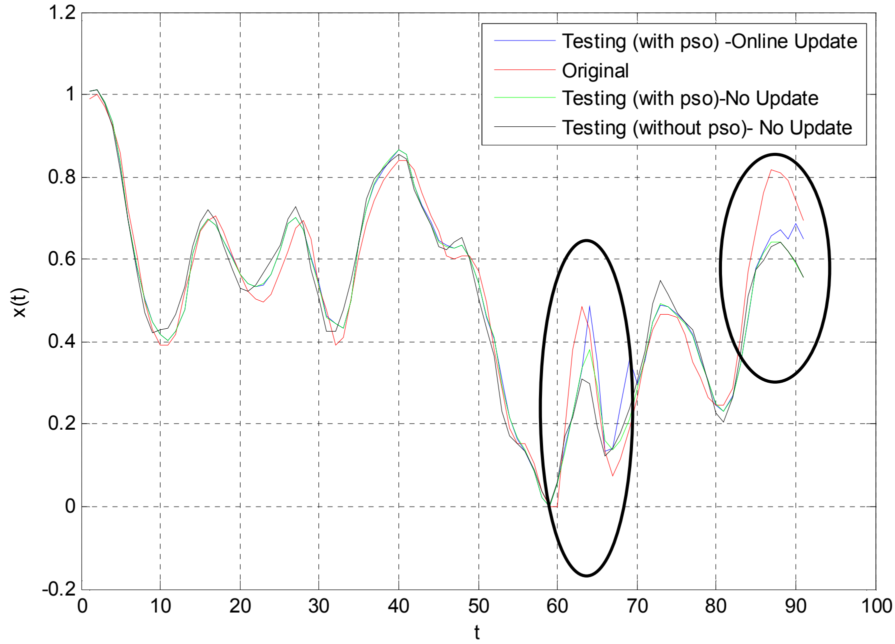


Figure 4-20: Gas Furnace Testing Results for DENFIS and MDENFIS (with and without online Updates)

During the testing stage the MDENFIS was continuously acquiring input-output data pairs when they become available to update the model, during which 3 new clusters were added in this example, hence better prediction accuracy is obtained. A comparison with other well-known models is listed in Table 4-8 below. MDENFIS with and without the online update capability outperformed all of them when the MSE error is compared.

Table 4-8 Gas Furnace Dataset 2 Testing Comparison

Method	# of Inputs ,Rules	MSE Error
ARMA (Kasabov 2007)	5 inputs, -	0.71
ANFIS (Kasabov 2007)	2 inputs, 25 rules	0.0073
FuNN (Kasabov 2007)	2 inputs, 7 rules	0.0051
HyFIS (Kasabov 2007)	2 inputs, 25 rules	0.0042
MDENFIS (without online update)	2 inputs, 11 rules	0.0033
MDENFIS (with online update)	2 inputs, 14 rules	0.0031

4.7 CASE STUDY No.5

4.7.1 Hypothesis

Ability to isolate and identify the source of a problem with high accuracy: It is expected that the model can isolate and identify the source of a problem using the automatic diagnostic part with high accuracy when compared with other methods.

4.7.2 Testing

The iris dataset 3 is used for this test. 150 inputs/output pairs were splitted equally into two datasets each with 75 inputs/output pairs, one for training the model and the other for testing it. Other tests using 150/150 training and testing inputs/output pairs and 120/30 training and testing inputs/output pairs were conducted.

4.7.3 Results and Discussion

The result of the first test (I) using 75 inputs/output pairs for training and 75 inputs/output pairs for testing is shown in Table 4-9 and is compared to test (II) (Swain et al. 2012). The results of test (II) were based on 500 iterations whereas the results of test (I) are based on one pass (1 iteration). The proposed model has two misclassified cases whereas test (II) misclassified 15 cases. Test (II) was repeated for 5000 iterations and results showed 5 misclassified cases. It's obvious that the processing time will short for test (I) with one pass compared to 5000 iterations with test (II) having in mind the accuracy isn't as good. The accuracy of the best results obtained with test (II) was 93.3% whereas the accuracy of test (I) was 97.3%. The results of test (I) are graphically shown in Figure 5-21 with two misclassified cases circled in red. The classes used in the figure are 1, 2 and 3 referring to Iris Setosa, Iris Virgnica and Iris Versicolor respectively.

The second and third testing results based on 150/150 and 120/30 inputs/output pairs for the training and testing are shown in figures 5-22 and 5-23. A 100% classification accuracy obtained with 150 inputs/output pairs for the training and testing with no misclassified cases and 96.66% classification accuracy with 1 misclassified case out of 30 cases obtained with 120/30 inputs/output pairs for the training and testing respectively.

Table 4-9 Iris Classification Using ECMcm and Neural Network Comparison

Class	Pairs #	Classified		Misclassified	
		I	II	I	II
Iris Setosa	25	25	24	0	1
Iris Virgnica	25	25	11	0	14
Iris Versicolor	25	23	25	2	0

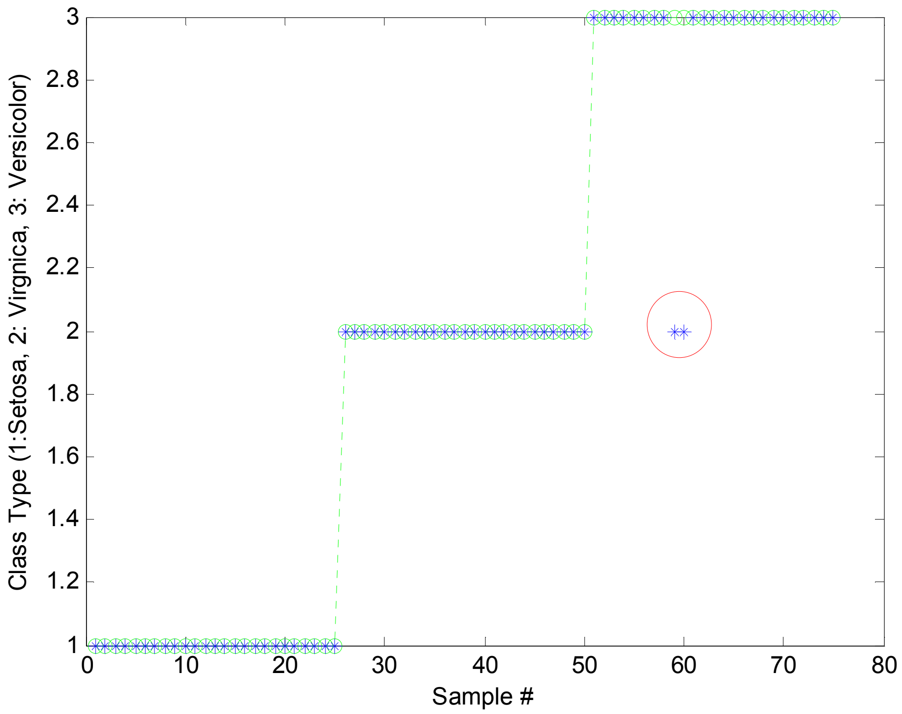


Figure 4-21 Iris Testing Results Showing Two Misclassified Cases

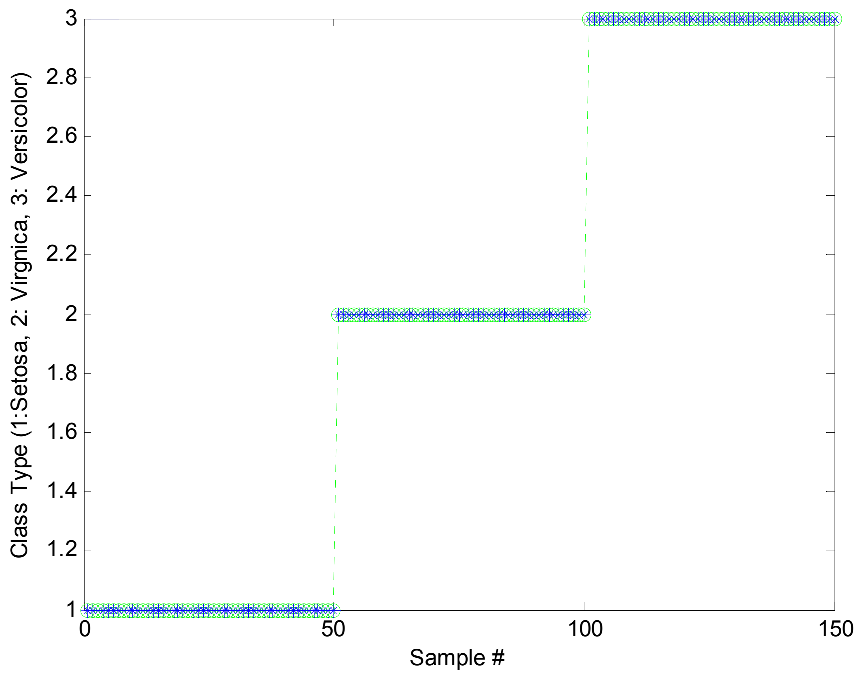


Figure 4-22 Iris Testing Results 150/150 inputs/output pairs (11 Clusters)

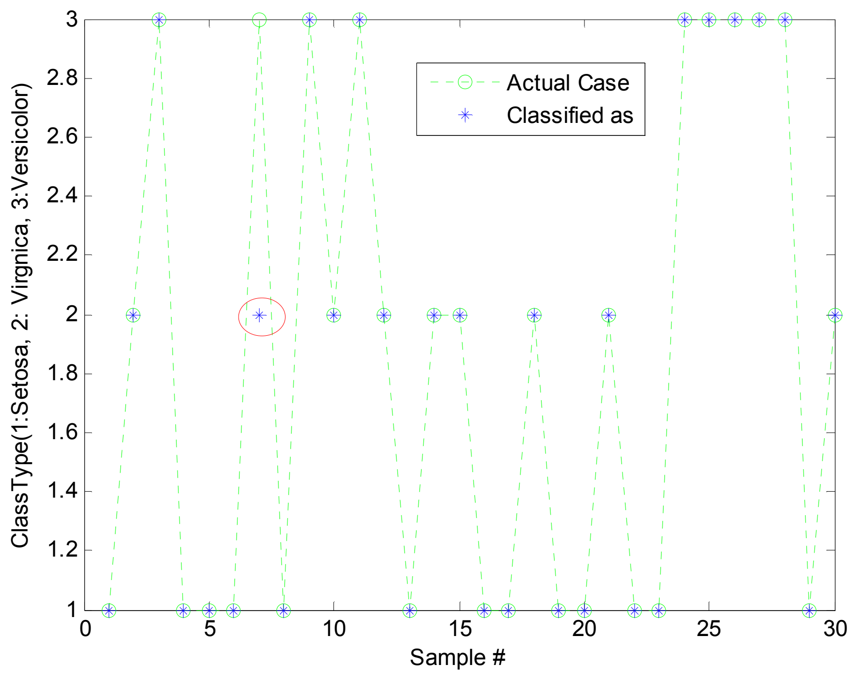


Figure 4-23 120/30 Testing inputs/output pairs (12 Clusters)

Table 4-10 Iris Testing Accuracy using Different Dataset Sizes

Method	**Accuracy**
Proposed Model 2 (150/150), 1 pass	100%
Proposed Model 3 (120/30), 1 pass	96.66%
Proposed Model 1 (75/75), 1 pass	97.30%
(Chen and Fang 2005) Proposed Model (150/150)	97.33 %
(Chen and Fang 2005) Proposed Model (120/30)	96.72 %
(Chen and Fang 2005) Proposed Model (75/75)	96.28 %
Dasarathy's method (150/150), see more details in (Chen and Fang 2005)	94.67%
Castro's method (120/30), 200 iterations, see more details in (Chen and Fang 2005)	96.60%
Hong-and-Lee's method (75/75), 200 iterations, see more details in (Chen and Fang 2005)	95.57%
Chang-and-Chen's method (75/75), 200 iterations, see more details in (Chen and Fang 2005)	96.07%

(Chen and Fang 2005) Presents a comparison table for different methods of classifying the Iris dataset using different training and testing dataset sizes. The results of the above proposed model is added on top of the table and are all presented here for comparison purposes in Table 4-10.

The highest accuracy achieved with our proposed model for 150/150 training/testing datasets with 100% accuracy of classification.

The lowest accuracy was 94.67% using Dasarathy's method (150/150) (Chen and Fang 2005). In general, the proposed model with different dataset sizes achieved either similar accuracy to other models with similar dataset sizes or better within 1-2 % higher.

4.8 SUMMARY

Five different case studies are presented in this chapter in order to proof the truth of a number of hypothesis claimed, by comparing the results of the proposed model with some of the other models available in the literature. Three benchmark datasets are used for testing the hypothesis. Four main improvements are obvious from the results:

1- Enhanced Testing Prediction Accuracy

2- Comparable Processing Time if not better

3- Ability to detect sudden changes in the process, hence, adding new rules during the testing stage.

4- Ability to isolate and identify the source of a problem with high accuracy

Chapter 5 will cover further model validation case studies however using field dataset obtained from a fouled centrifugal compressor in an offshore installation. Chapter 5 will also present a fully integrated Condition Based Maintenance Framework including the early detection of any deviation from the normal operating envelope. Diagnostics, short term and long term prediction (prognostics).

5 MODEL VALIDATION THROUGH OFFSHORE OIL AND GAS CASE STUDY

The proposed model was validated using benchmark datasets used by other researchers in the fields of Artificial Intelligence, Pattern Recognition, Time Series Prediction, etc. However, one of the main objectives of this research is to develop a model that is easy and practical to be deployed in real field applications. The condition monitoring and performance data of a low pressure centrifugal compressor installed offshore are used to train the proposed model and test its capabilities. Section 5.1 gives an introduction to compressors types with more emphasis on centrifugal compressors and their role in the offshore industry. Section 5.2 discusses the condition monitoring requirements for centrifugal compressors and section 5.3 gives an overview of the performance calculations required to obtain the calculated tags used within the proposed model. Section 5.4 gives a brief description of centrifugal compressors fouling types, sources and methods of detection. This case will be used to detect the initiation of fouling and its progression short term and long term (prognostics). The assessment results of the integrated condition monitoring and prognostics proposed model are presented in section 5.5.

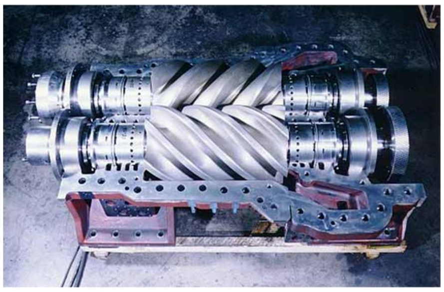
5.1 Introduction

There are two basic types of compressors: dynamic compressors and positive displacement compressors each of which uses different mechanism to compress the gas. With positive displacement compressors, the flow supply is intermittent whereas a continuous flow supply is provided for dynamic compressors. A Positive displacement compressor entraps the gas in a compartment then reduces its volume increasing the gas pressure, examples are rotary screw compressors, reciprocating (piston) compressors, etc. With dynamic compressors, the energy generated during rotation of the rotor is transferred to the gas. Centrifugal and axial compressors are examples of dynamic compressors. The Main differences between positive displacement and dynamics compressors are shown in Table 5-1:

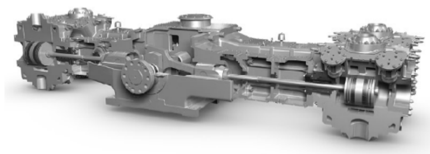
Table 5-1 Positive Displacement and Dynamic Compressors Main Characteristics

Characteristic	Positive Displacement Compressor	Dynamic Compressor
Flow Supply	Intermittent	Continuous
Volume	Constant, low flow rates	Variable, higher flow rate
Energy (Head)	Variable	Constant for a certain flow
Gas Characteristics	Independent	Dependent
Maintenance Requirements	Low	High

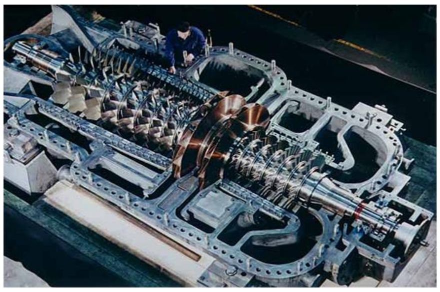
Figures (5-1-a,b, c,d) show respectively photos for rotary screw, reciprocating, axial and centrifugal compressors.



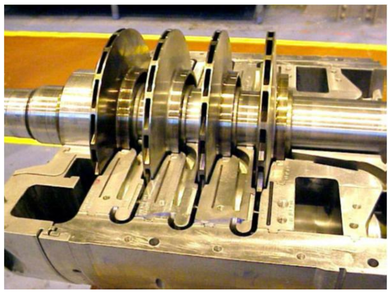
a. Rotary Screw Compressor



b. Reciprocating Compressor



c. Axial Compressor



d. Centrifugal Compressor

Figure 5-1 Examples of Different Types of Compressors

Centrifugal compressors consist of one or more impellers rotating at high speed during operation. The impellers increase the speed and pressure of the gas which flows in a radial direction in the impeller.

In any offshore installation the two main hydrocarbons dealt with are natural gas and crude oil. The two hydrocarbons in addition to water are produced as a mixture from the production wells. The flow is controlled through a production choke and is directed towards the separation system. In order to enhance the separation process the liquid is going through two processes upstream of the separation system, namely heating and chemicals addition. The separation system, which may consist of one or more separators, separates the fluid into oil, gas and produced water, each directed into different path for further processing. The gas recovered from the separation system is then cooled allowing the liquids entrained to condense out and rest at the base of the scrubber. The gas recovered flows towards the gas compression system. It is important to ensure that liquid isn't carried over with the gas to the gas compressor as this will cause corrosion, potential ice and hydrate formation. The process flow is shown in Figure 5-2 below (Level 1 & 2).

Within the gas compression system at least one stage of compression is in place. Multistage compression could be in the form of a number of stages within same casing or in individual casings. The compressor is normally driven by an electric motor or gas turbine via power transmission units (Gearbox and couplings).

Supporting the operation of the compressor are a number of auxiliary systems including: lubrication system, shaft seal system, control and monitoring system and other miscellaneous equipment like purge air system. The gas produced from the gas compression system is then used for many different purposes including:

1. Exported through subsea pipelines for sale.
2. Injected back to the wells to help maintain pressure in the reservoir.
3. Injected back to the wells to help them flow, i.e. as gas lift.
4. Used as a fuel and Flaring

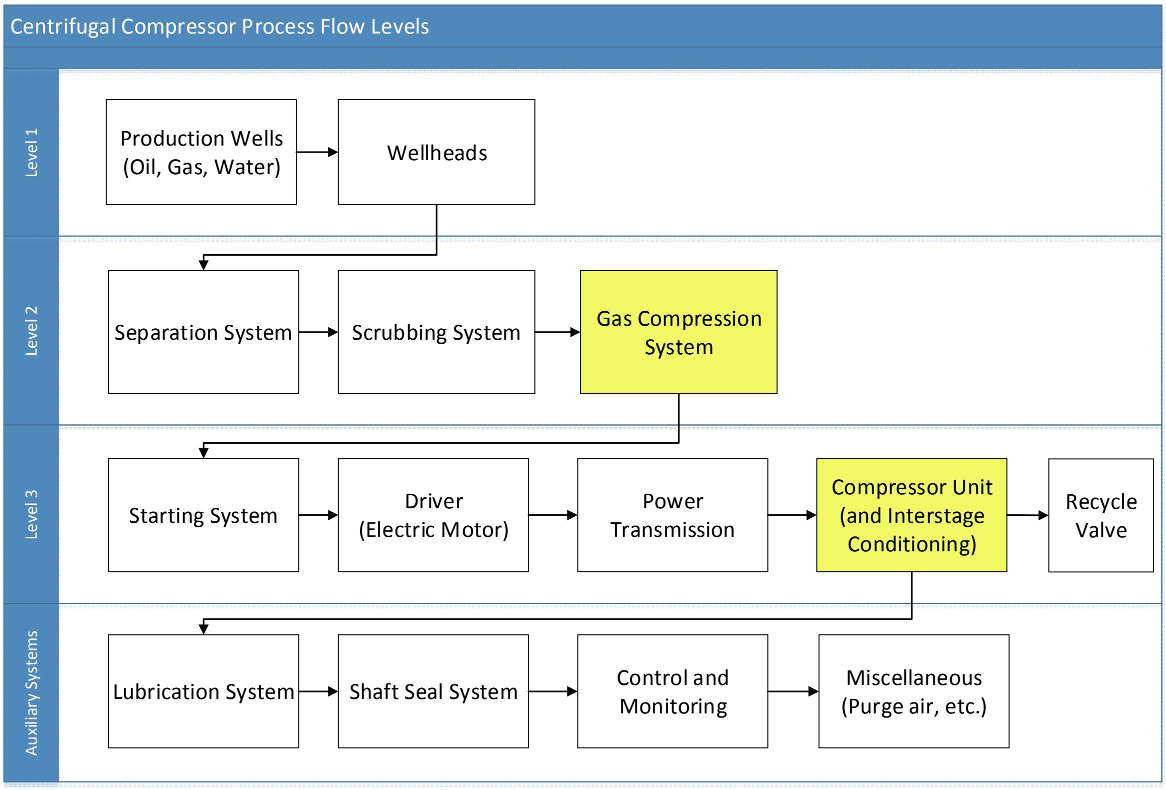


Figure 5-2 Centrifugal Compressor Process Flow Levels

5.2 Condition Monitoring of Centrifugal Compressors

Figure 5-3 shows the major mechanical components of a centrifugal compressor. Those can be summarised as the following:

1. Casing: vertical or horizontal split. Provides pressure containment and houses both rotating and static elements of the compressor.
2. Rotor including impellers and shaft
3. Radial Bearings: Support the rotor weight.
4. Thrust Bearings: Support any excess thrust load not supported by the impellers design and/or the balance piston.
5. Balance Piston: Counteracts the net thrust load developed due to the existing pressure difference in the direction of the compressor inlet.
6. Seals: Prevent compressed gas leakage or oil and/ or air getting inside the compressor and getting mixed with the gas.

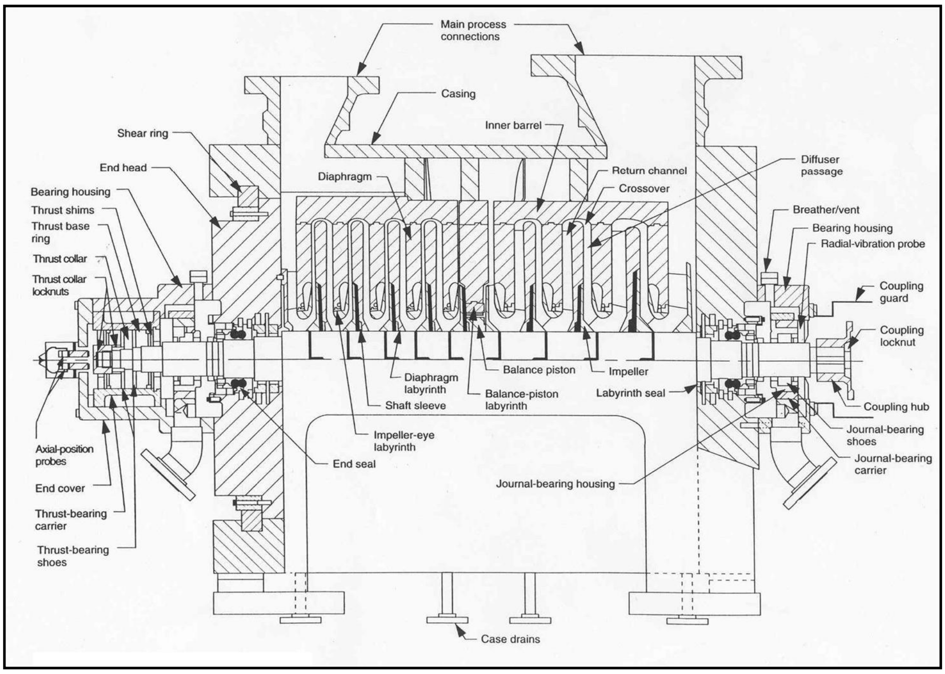


Figure 5-3 Centrifugal Compressor Main Components

To support the operation of the gas compression train a number of auxiliary systems are used including the lubrication system, shaft seal system, control and monitoring and other miscellaneous equipment. Figure 5-4 shows a list of common maintainable items for a generic centrifugal compressor train derived from ISO 14224:2006 (ISO 14224 2006).

In order to evaluate the overall health of the centrifugal compressor train the condition of each of these maintainable items should be monitored. Figure 5-5 is an example of which parameters need to be monitored for each maintainable item linked to the failure mode(s) they are most sensitive to. This isn't in anyway an exhausted list and is only used here as an example.

Equipment Class	Compressor					
Subunit	Power Transmission	Compressor	Control and Monitoring	Lubrication System	Shaft Seal System	Miscellaneous
Maintainable item	• Gearbox • Bearings • Couplings • Lubrication • Seals	• Casing • Rotor • Radial Bearings • Thrust Bearing • Balance Piston • Seals • Antisurge System	• Pressure and Temperature Transmitters • Control Valves • Pressure and Temperature Controllers • Corrosion Monitoring • Cables and • Junction boxes • Internal power • Supply • Piping	• Reservoir • Pumps • Coolers • Filters • Piping • Valves • Lube oil	• Reservoir • Pumps • Coolers • Filters • Piping • Valves • Seal oil • Dry gas seal	• Base Frame • Piping and supports • Valves • Coolers • Silencers • Purge air • Flange joints

Figure 5-4 Centrifugal Compressor Train Maintainable Items as per ISO 14224

Maintainable Item	Failure Mode Description	Polytropic Efficiency (%)	Bearing Temp. (Deg C)	Bearing Vib. (Micron p-p)	Suction Pressure (bar a)	Suction Temp. (Deg C(	Discharge Pressure (bar a)	Discharge Temp. (Deg C)	Flow Rate (m3/hr)	Recycle Valve Position (%)	Lube Oil Level	Lube Oil Temp. (Deg C)	Lube Oil Pressure (bara)	Lube Oil Filter DP (bara)	Rotor Axial Position	Seal Gas Pressure (bara)	Seal Gas Filter Dp (bara)	Seal Gas Flow	Lube Oil Analysis
		Matrix mapping relationships between Failure Modes and Observed Effects																	
Radial Bearing	Damage due to lack of Lubrication		●	●							●								
Radial Bearing	Damage due to (brinelling, fretting wear etc.)		●	●															●
Thrust Bearing	Damage due to excessive vibration		●	●											●				
Rotor	Internals worn / damaged (impellor, O rings etc.)	●		●			●	●	●										
Rotor	Deposits buildup/Fouling	●	●	●			●	●	●										
Rotor	Loose parts			●															
Antisurge System	Surging						●		●	●					●				
Pressure Transmitter	Abnormal instrument reading (suction)				●														
Filter	Lube Oil Filter Blockage												●	●					
Dry Gas Seal	Failure to contain fluid															●	●	●	
Valve	Lube Oil PCV malfunction												●						
Piping and Support	Piping strain		●	●															
Baseframe	Soft foot		●	●															

Figure 5-5 Generic Centrifugal Compressor Matrix Mapping Relationships between Failure Modes and Observed Effects

With the maintainable items list, their functions and failure modes, their most sensitive condition monitoring parameters are all defined, the remaining task is to define data sources and frequency of taking measurements. A failure mode with short lead time to failure requires a permanent online data acquisition system, whereas a failure mode with longer lead time to failure may be managed through the use of an offline portable data collector.

The lead time to failure and criticality classification of equipment taking into consideration the risk involved (safety, environmental and financial aspects) are the two main aspects in deciding the level of condition monitoring. Centrifugal Compressors on an offshore application are essential non-spared pieces of equipment which require an online condition monitoring, they are usually classified as critical pieces of equipment.

The main objective is to monitor the mechanical and performance behaviour of the compressor to enable optimal equipment operation and reduce any unplanned maintenance activities hence enhancing the reliability and availability of the compressor. The parameters list in Figure 5-5 are the minimum list of parameters to be included as part of any condition monitoring program for centrifugal compressors. The following can also be added as necessary:

1- Phase Angle and speed (for compressors with variable speed drives)

2- Gas Composition of the process gas (for accurate performance calculations)

3- Separator Liquid Level as an indication of liquid carry over

4- Electric motor current and/or power (if the compressor is motor driven)

5- Suction strainer differential pressure if one is installed.

Condition Monitoring (CM) is measuring and trending of the condition and performance indicators of equipment so that the scope and timing of maintenance interventions (overhaul, repair and/or design and process changes) are known in time to minimise the impact on production and to ensure that resources are available to carry out maintenance.

Figure 5-6 shows an example failure mode PF curve. P refers to the point at which a potential failure is detected, and F is the point at which a functional failure occurs. There is an area preceding the potential failure point, tagged as the normal operating envelope area, during which the machine will be running within its normal operating envelope, obviously with variations in the operating and environmental conditions, the trend will not be constant however it will vary within the dashed blue lines (normal operating envelope boundaries).

Any deviations from the normal operating envelope boundaries means that a fault could potentially develop. Detecting this fault at an early stage will allow a timely maintenance intervention to be taken in order to minimise the impact on production and to ensure that resources and materials are available to carry out the maintenance work. Various techniques are being developed in the literature and are discussed in Chapter 2 that serve the purpose of early warning detection, diagnostics and fault finding but ultimately the main goal of condition monitoring and condition based maintenance is to predict the remaining useful life of rotating equipment to enable proper planning for resources and materials ensuring minimum impact on production.

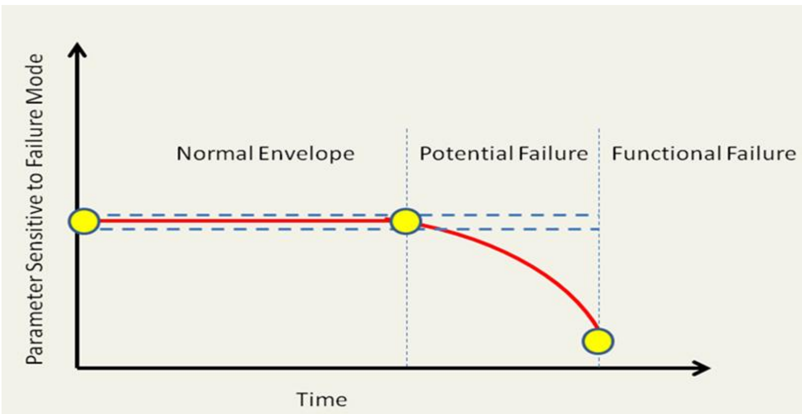


Figure 5-6 PF Curve for A specific Failure Mode

5.3 Centrifugal Compressor Performance Calculations

Compressor performance is an important health indicator to evaluate the current condition and reliability of the compressor. The calculations for the compressor performance are based on gas properties and thermodynamics of compressors. It is imperative when doing the calculations to obtain good quality data from the field as any instrumentation issues will have significant impact on the quality of the data results. Tests are normally done in accordance with ASME PTC 10 (ASME PTC-10 1997) and considering real gas behaviour as opposed to ideal gas law assumptions.

Using ideal gas equations will give approximate results which in some cases can be appropriate especially with single component gases at low pressures however with gas mixtures, especially at elevated pressures, ideal gas laws are no longer applicable and more advanced techniques are required. The introduction of the compressibility factor as an indication of deviation from the ideal behaviour into the ideal gas equation becomes vital.

i.e. the ideal gas equation becomes:

$$PV_m = ZR_oT \tag{5-1}$$

Where Z is the compressibility factor;

For hydrocarbon gas mixtures, utilising the following Equations of State give better results when compared with the graphical method like Mollier diagrams (Rajput 2009) and using thermodynamic tables: Lee Kesler (Lee and Kesler 1975) and Benedict-Webb-Rubin (BWR) (Benedict et al. 1940).

The following parameters are required to complete the compressor performance calculations using a Polytropic process:

1- Suction and Discharge Pressures in bara (P_{suc}, P_{dis})

2- Suction and Discharge Temperatures in K (T_{suc}, T_{dis})

3- Gas flow rate in mmscfd (Q)

4- Gas mixture composition

5- Rotational speed (for compressors running at variable speeds, ω)

6- Driver load (I)

The original equipment manufacturer is required to supply the design performance curves represented in several ways:

1- Discharge Pressure Versus Flow Rate at constant inlet conditions

2- Differential Pressure Versus Flow Rate at constant inlet conditions

3- Polytropic Efficiency Versus Flow Rate at constant inlet conditions

4- Polytropic Head Versus Flow Rate at constant inlet conditions

Due to changes in the inlet conditions (suction pressure, temperature and gas composition) the first two curves might not be useful as this will have significant effect on the discharge pressure and differential pressure. The curves for the Polytropic efficiency and head do not change significantly with moderate changes in the inlet conditions. Three important regions in the curves are normally defined:

A best efficiency point (BEP) and an envelope region around it showing the best operating point and the normal operating envelope for the compressor. The area to the left of the curve at low flow rates and high discharge pressures, starting from a point (for single speed compressors) or a line (for variable speed compressors). This point (or line) defines the minimum flow below which reversal flow will occur, a condition called surge which is very destructive to the compressor internals. Normally this is picked up by the control system, and a command is sent to the antisurge control valve to open allowing some of the discharge flow to divert its direction towards the suction of the compressor, shifting the operating point of the compressor away from the surge region.

The area to the right of the curve at high flow rates and low discharge pressures is called stonewall or choke, beyond which the compressor cant increase its flow any further. The main effect of operating the compressor in a choke condition is at the driver side as overloading the driver (if electrically driven) will increase the windings temperature and can ultimately cause a motor burn out. However, if the balance piston is properly size providing the required thrust balance minimal effect will be seen at the compressor side.

The following curves are some examples of typical curves, Figures 4-7 to 9 below.

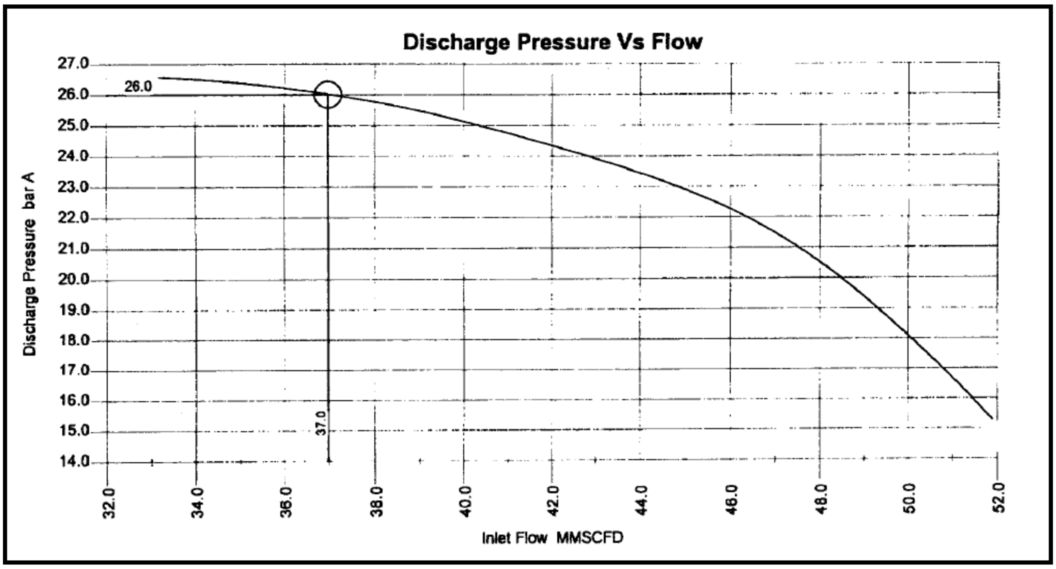


Figure 5-7 Example Discharge Pressure versus Flow Rate Curve

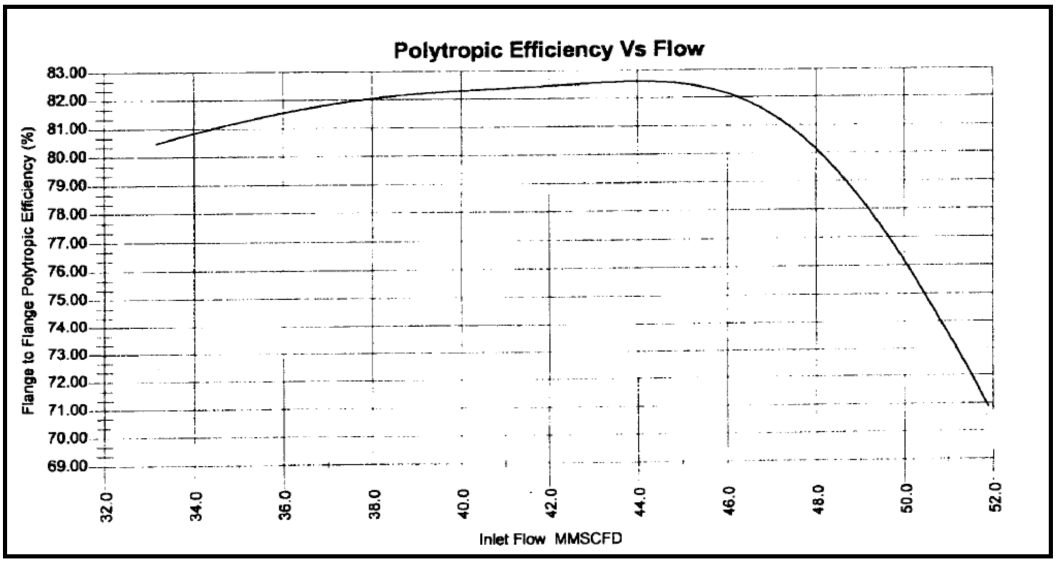


Figure 5-8 Example Polytropic Efficiency versus Flow Rate Curve

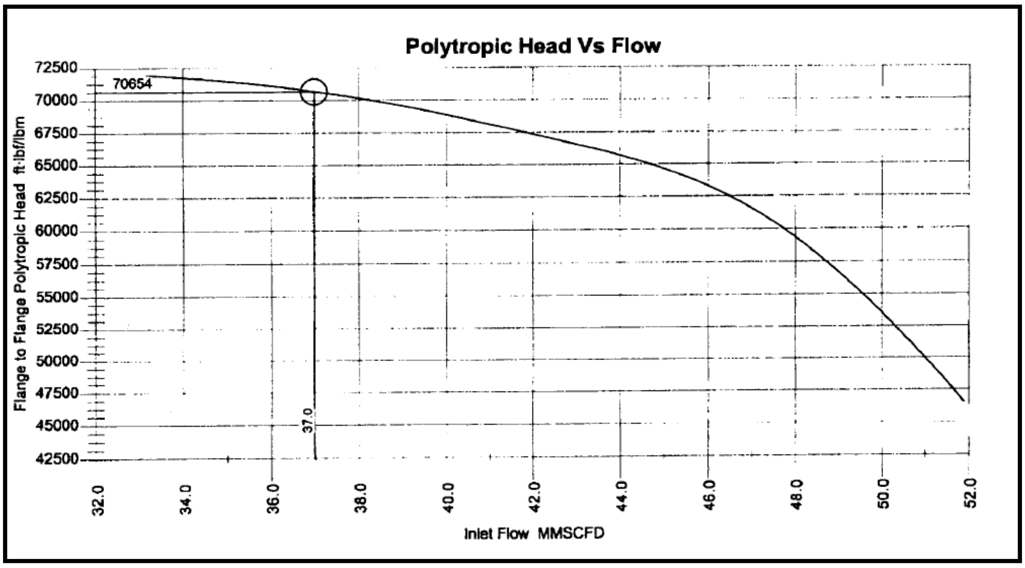


Figure 5-9 Example Polytropic Head versus Flow Rate Curve

The equations shown below specify the requirements for compressor performance calculation in a simplified way, however when doing the actual calculations, ASME PTC 10 procedures (ASME PTC-10 1997) should be followed:

1- Gas Mixture Molecular Weight Calculation

$$MW = \sum_{1}^{i} MW_i * Mole\ \ Fraction_i \qquad \textbf{(5-2)}$$

Where MW_i and $Mole\,\mathrm{Fraction}_i$ are the molecular weight and Mole fraction of the ith gas component

2- Gas Mixture Critical Pressure

$$P_c = \sum_{1}^{i} P_{c\,i} * Mole\ Fraction_i \qquad \textbf{(5-3)}$$

Where P_{ci} and $Mole\ Fraction_i$ are the critical pressure and Mole fraction of the ith gas component;

3- Gas Mixture Critical Temperature

$$T_c = \sum_{1}^{i} T_{c\,i} * Mole\ Fraction_i \qquad \textbf{(5-4)}$$

Where T_{ci} and $Mole\ Fraction_i$ are the critical temperature and Mole fraction of the ith gas component respectively;

4- Gas Mixture Specific Heat as function of Temperature

$$C_p = A + BT + CT^2 + DT^3 + ET^4 \tag{5-5}$$

Where A, B, C, D and E are the correlation constants obtained from thermodynamics tables (Coker 2010);

5- Isentropic and Polytropic exponents

$$k = \frac{C_p}{C_p - R_o} \tag{5-6}$$

Where k is the isentropic exponent;

$$\frac{n}{n-1} = \frac{\ln\left(\frac{P_{dis}}{P_{suc}}\right)}{\ln\left(\frac{T_{dis}}{T_{suc}}\right)} \tag{5-7}$$

Where $\frac{n}{n-1}$ is the Polytropic exponent.

6- Ideal Gas Polytropic Efficiency

$$\eta_p = \frac{\frac{n}{n-1}}{\frac{k}{k-1}} \tag{5-8}$$

7- Ideal Gas Polytropic Head

$$H_p = \frac{8314 * Z_{avg} * T_{suc}}{MW\left(\frac{n-1}{n}\right)}\left[\left(\frac{P_{dis}}{P_{suc}}\right)^{\frac{n-1}{n}} - 1\right] \tag{5-9}$$

5.4 Centrifugal Compressor Fouling

5.4.1 Introduction

Fouling is the accumulation of undesired deposits and droplets throughout the rotor assembly of the compressor including: Inlet volute, impellers, diffusers, balancing piston, balance line, discharge volute and inner seals, etc. which seriously deteriorate the performance of the machine and limit its capacity.

Performance loss due to fouling can be more pronounced on low flow compressors since the flow passages are very narrow (ex. Diffuser width can be just a few millimetres wide), and accumulation of particles can cause significant restriction for the flow and pressure losses.

Figure 5-10 Fouled Centrifugal Compressor Rotor

5.4.2 Fouling Types and Causes

There are various types of fouling (Mostafa 2011) including:

1. Particulate Fouling: Accumulation of particles
2. Crystallization or Precipitation Fouling: e.g Crystallization of dissolved salt.
3. Chemical Reaction Fouling: e.g. Polymerization and Coking
4. Corrosion Fouling: Chemical reaction between the gas mixture and the surface to produce corrosion material which accumulates on the surface or transfers with the gas and accumulate on different areas fouling them.
5. Solidification or Freezing Fouling

Of the above, the two types commonly found in turbomachinery are particulate fouling and corrosion fouling (Drury and Cockrell). However, from experience a third type can be added to the list especially with centrifugal compressors handling hydrocarbons which is Chemical reaction fouling specifically polymerisation and coking. The three fouling types will be discussed in some details below:

Particulate Fouling: is the build-up of heavy dirt/particles entered to the compressor with the gas, which can clog the internals of the compressor. This material build-up roughen the aerodynamic surfaces limiting the capacity of the compressor, increasing the pressure losses and decreasing the overall thermal efficiency of the compressor (Hanlon 2001). Main factors affecting this type of fouling are temperature, velocity and gas composition including concentration of suspended particles (Mostafa 2011). One example is iron carbonate deposits from corrosion of upstream pipe work systems.

Figure 5-11 Fouled 1st Stage Discharge- Particulate Fouling

Polymerisation and coking: Polymerisation involves a chemical reaction where sticky hydrocarbon aerosols bond to the compressor base metal causing deterioration to the performance of the compressor and other problems that will be discussed in the next section. Main factors affecting this type of fouling are temperature (above 90 Deg C) (Hanlon 2001), pressure (pressure level proportional to the fouling limit), surface finish (smoother surfaces are less susceptible to fouling) and finally the gas composition and concentration of hydrocarbon aerosols (Hanlon 2001). Coking is also a chemical reaction whereby thermal degradation of one of the gas mixture components occurs, resulting in carbonaceous deposits on the internals of the compressor.

Figure 5-12 Coking at the stage inlet

Corrosion Fouling: Involves electrochemical reaction where significant moisture in the gas presents containing salts and acids. Condensed water coming in contact with the compressor internals causing aqueous corrosion. Corrosion can be localised or uniform over the entire internals of compressor.

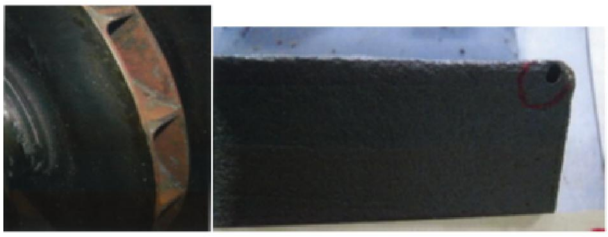


Figure 5-13 Impeller Corrosion and Return Bend Pitting

5.4.3 Fouling Effects on Centrifugal Compressors

Fouling has many adverse effects on centrifugal compressors including but not limited to:

1. Reduced Operating Efficiency due to increased frictional losses and internal recirculation, (Turner), (Meher-Homji 1989), (Snider 2006), (Gresh 2001), (Song 2005) and (Hanlon 2001).
2. Reduced Head, (Turner).

3. Reduce pressure ratio due to the following process: particles accumulated on the impellers roughen the surface leading to flow restriction causing a reduction in the flow rate and decrease the pressure ratio, (Song 2005).
4. Reduced Maximum flow rate, (Turner) and (Snider 2006).
5. Reduced Component Reliability and Life.
6. Build-up of Material Causing Unbalance and High Vibration long term causing fatigue, (Turner) and (Meher-Homji 1989).
7. Reduce radial and axial clearances causing severe wear damage to the inter-stage labyrinths seals and impellers (Snider 2006) and (Hanlon 2001).
8. Failure to the thrust bearing due to surging when fouling is in the balance piston labyrinths and balance line, (Turner), (Meher-Homji 1989).
9. Impellers Integrity effects, (Meher-Homji 1989).
10. Erosion, Corrosion and diffuser passage blockage, (Meher-Homji 1989) and (Hanlon 2001).
11. Negative financial effects due to restricted maximum throughput, increased operational cost and potentially requiring unscheduled shutdown, (Snider 2006).

5.4.4 Monitoring Fouling

Measuring the build-up rate of fouling is a hard task given that the process is dependent on so many different issues, including the gas composition in processes where variation of gas composition is just normal. This isn't always predictable, however given that the fouling effect on centrifugal compressors is known, correlated parameters are monitored to give an indication about the deterioration rate. Combining this data with overhauls as found fouling thickness measurements for stable processes can help predicting the build-up rate of fouling and the severity of the situation. The following are a list of the main compressor condition monitoring parameters that can be trended to measure the effect of fouling on the compressor operation:

1. Polytropic efficiency
2. Polytropic Head

3. Polytropic exponent
4. Pressure Ratio
5. 1X vibration and Phase in the radial direction
6. Thrust Position and Vibration
7. Thrust Temperature Difference between active and inactive sides
8. Intercooler pressure drop
9. Balance line differential pressure

Figure 5-14 shows an example of Polytropic efficiency data versus flow rate calculated from process data historian over 8 years showing clear degradation in the performance from 79% at 37.5 mmscfd in 2006 to 63% in 2012/2013. The data is calculated using Peng Robinson EOS and are coloured by year. The compressor was taken offline in 2013 for a bundle change and Figure 5-7 shows the fouled old bundle.

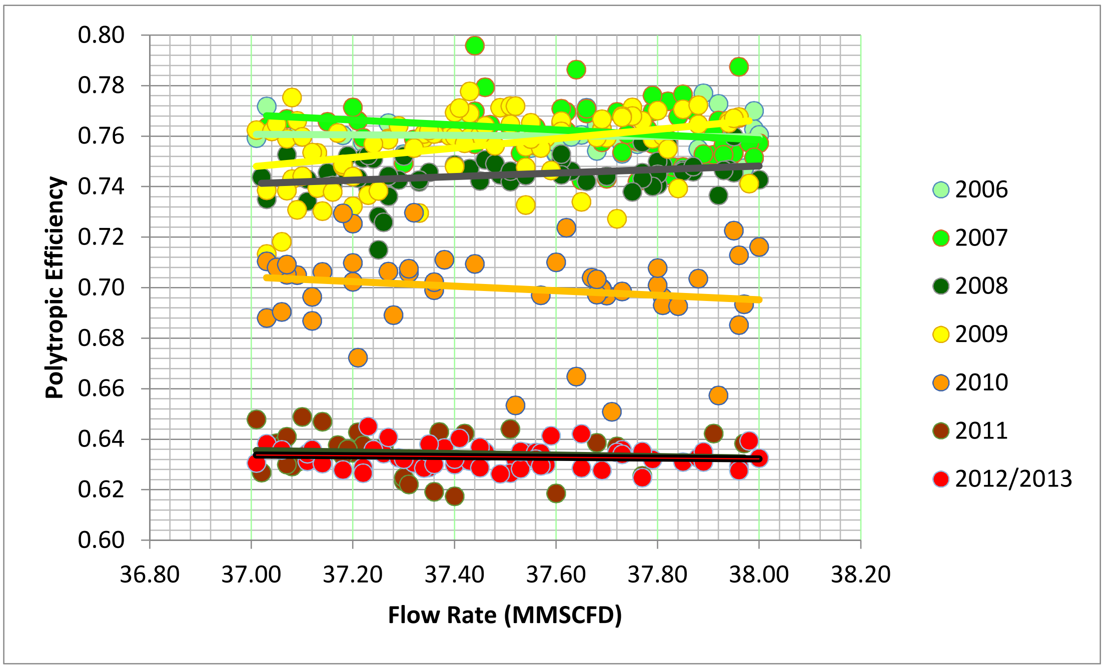


Figure 5-14 Centrifugal Compressor Polytropic Efficiency versus Flow Rate over 8 years

5.5 Proposed Integrated Condition Monitoring and Prognostics Model Simulations

5.5.1 Introduction

This section presents a number of case studies to apply the proposed model in Chapter 3 on field data. A centrifugal compressor fouling case study was used for this purpose due to data availability and the fact that in this specific case fouling degradation was progressive with clear correlation with process data. The model currently allows interaction with subject matter experts (SMEs) as compared to a black box by selecting the appropriate inputs with high correlation to the output of interest. This can be an input for example from a FMEA/FMECA study. The accuracy of the prediction is highly dependent on the selected inputs to the model. Including inputs with low or no correlation might have an adverse effect on the model's performance and accuracy.

However due to the limited data available about more than one faulty condition to test the automated diagnostics algorithm, field data for another piece of equipment (Transformer) was used for this test. The first case study tests the capabilities of the model within the normal operating envelope region. Case study 2 covers the fault classification and diagnostics part using transformer gas concentrations dataset. Case study 3 presents the short term and long term prediction for the compressor fouling deterioration rate using relative efficiency drop as a health indicator.

5.5.2 Case Study 1: Normal Operating Envelope Monitoring

This section presents a new technique with simulations to allow early equipment failure detection. The following techniques are used for exactly the same purpose:

- Multivariate State Estimation Technique (MSET) (Herzog et al. 1998).
- Support Vector Machines (Chang et al. 2003).
- Principle Component Analysis (PCA) (Kano et al. 2000)
- Similarity Based Modelling (Wegerich 2005).

- Generalized Regression Neural Network (GRNN) (Specht 1991).

A comparison was made between the above techniques excluding the support vector machines in (Wegerich 2006) using the following performance metrics: Robustness, Spillover and Error. The comparison showed very close results in terms of the prediction error however the robustness and spillover metrics varied significantly between the different techniques with the similarity based modelling technique overperforming with two times and four times as good in terms of robustness and spillover respectively (Wegerich 2006).

This case study presents experimental testing for the proposed model using field data taken from a multistage centrifugal compressor, in addition to performance testing the model for robustness, spillover and error metrics.

The first step in creating a model is to draw some boundaries around the equipment and any auxiliaries included within the model. ISO 14224:2006 (ISO 14224 2006) has a good reference for rotating and static equipment boundaries that can be used here. Example is shown in Figure 5-15 for a compressor:

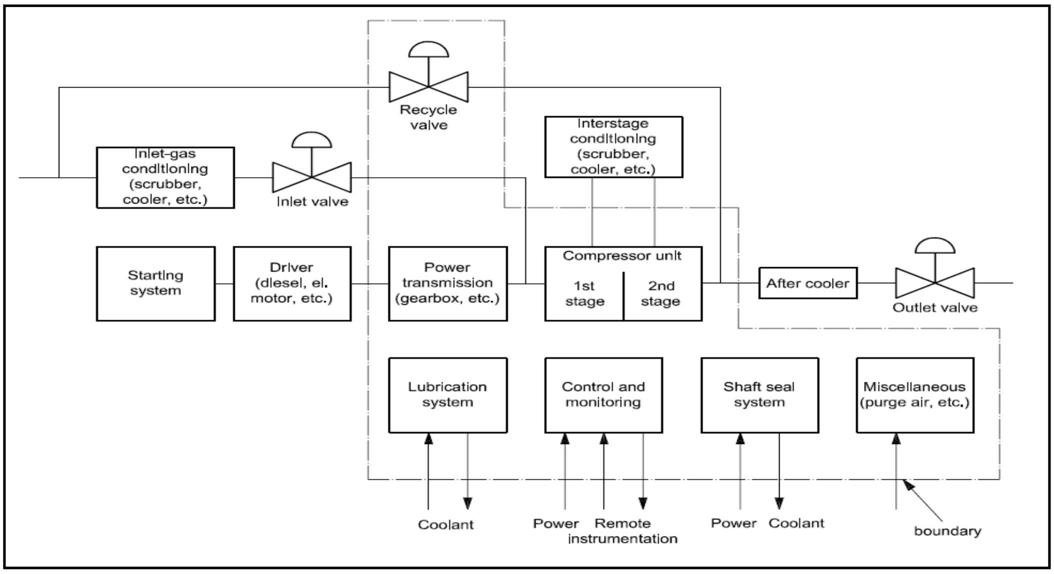


Figure 5-15: Boundary Definition- Compressors (ISO 14224:2006)

To simplify things, the boundaries are created around the compressor unit only and the model will cover the following performance parameters and some supporting operating parameters as listed in Table 5-2:

Table 5-2: Performance Model Parameters

#	Parameter Description	Parameter Unit
1	Suction Pressure	Bara
2	Suction Temperature	K
3	Discharge Pressure	Bara
4	Discharge Temperature	K
5	Flow Rate	MMSCFD
6	Polytropic Efficiency	%
7	Polytropic Head	%
8	% Relative Efficiency	%
9	% Relative Head	%
10	Temperature Ratio	-
11	Pressure Ratio	-
12	Polytropic Exponent	-

Although only 12 parameters are used here, the model has no limitations in terms of the number of parameters used. The data used is extracted from a process data historian for 6 years, however as long as the data contains variable operating conditions within the normal operating envelope less amount of data can be used for training the model as small as two weeks. The model has the capability of adding new ideal states (clusters) while in operation automatically if these new ideal states were not learned by the model during the learning stage, however this will only be done under strict conditions as otherwise a new state which might be an indication of a fault region (new cluster) is considered ideal and added to the normal operating envelope matrix. The extracted data is cleaned from the following:

- Zero speed/amp data (indicating that machine isn't running)
- Bad tags
- Any abnormal behaviour leaving only data within normal operating envelope.

ECM is then used for data reduction to remove any similarities between the different states due to repeated measurements at the same operating state leaving only a few number of examples representing the variations within the normal operating envelope, namely normal operating envelope matrix. This is followed by testing the model online using new data not used in training the model. Figure 5-16 shows the Polytropic efficiency data calculated and compared with the OEM design curve. The data is coloured based on time from dark blue (old data) to red (latest data). The Polytropic efficiency was dropping with time due to fouling and the whole curve was moving to the left.

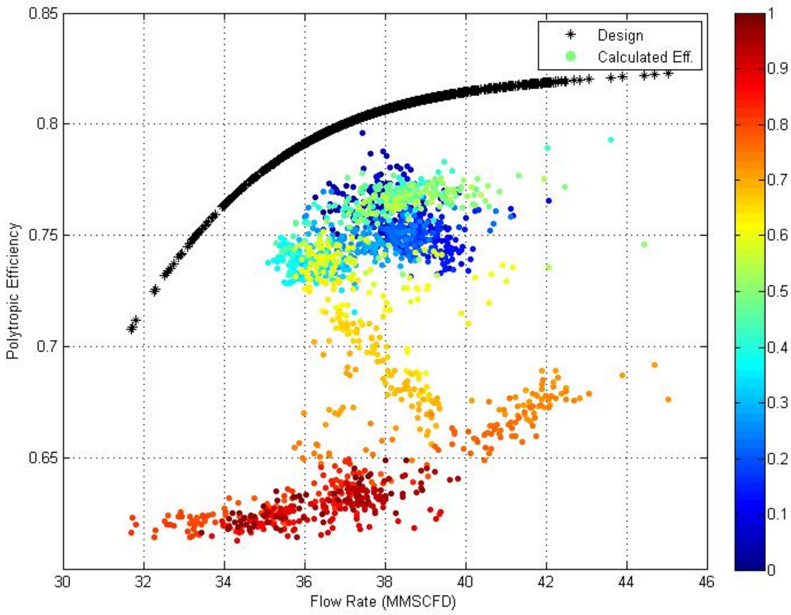


Figure 5-16: Polytropic Efficiency Map

Similarly, Figure 5-17 shows the Polytropic head data. After cleaning the data, Figure 5-18 shows the remaining data for the Polytropic efficiency that was used to train the model and obtain the normal operating envelope matrix alongside the other performance parameters collected and/or calculated at the same time instances. This equates to <7% drop in the efficiency as shown in Figure 5-19. The efficiency drop is calculated as the ratio between the actual Polytropic efficiency and the design Polytropic efficiency at a specific flow rate. The ideal flow rate, suction and discharge pressure and temperature trends are shown in Figure 5-20.

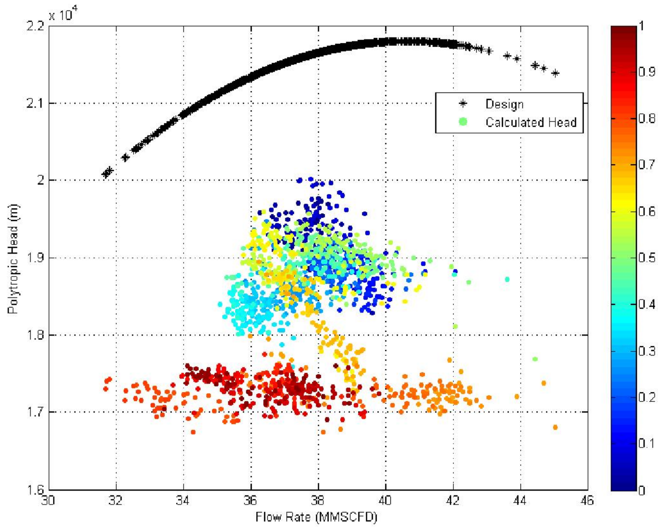


Figure 5-17: Polytropic Head Map

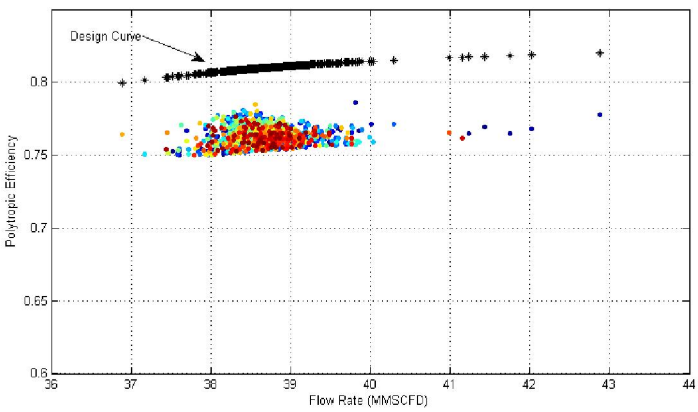


Figure 5-18: Polytropic Efficiency Trend over time (Ideal)

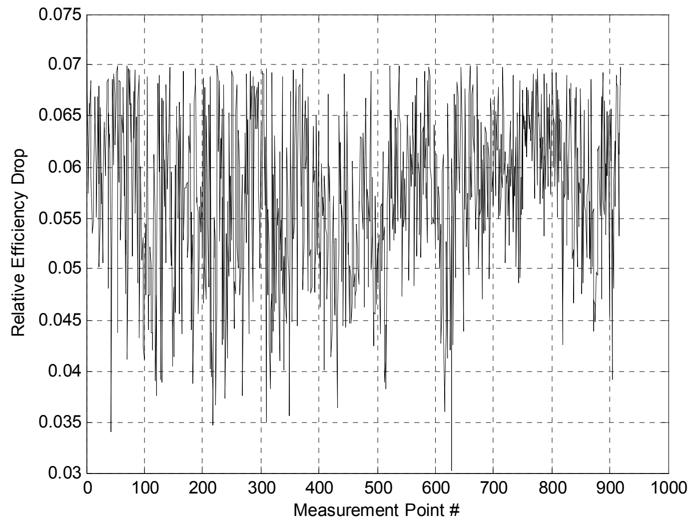


Figure 5-19: Relative Efficiency Drop (for the ideal model creation)

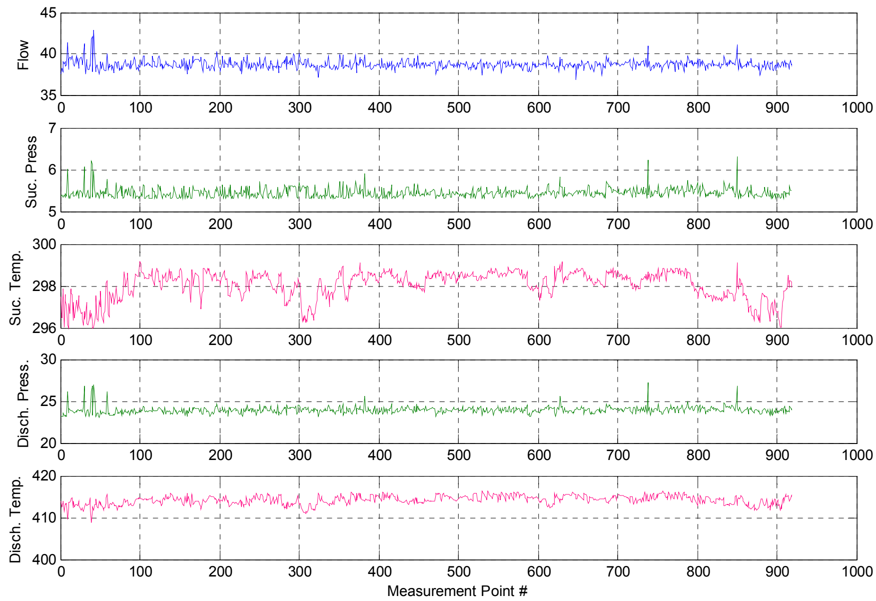


Figure 5-20: Ideal Performance Data Used for Model Training

As shown in Figure 5-21, 919 data pairs were used for each parameter. After clustering the data using ECM only 658 unique data pairs (normal operating envelope matrix) to be used during the testing stage are left. Figures 5-21, 5-22 and 5-23 shows the ideal data predicted by the model and the actual data. The two curves in each plot are identical indicating a good quality model.

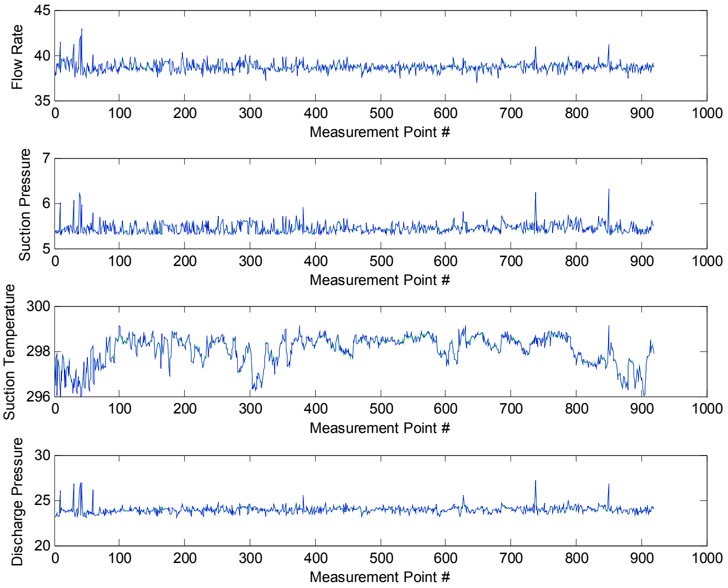


Figure 5-21: Ideal Performance Parameters Prediction 1

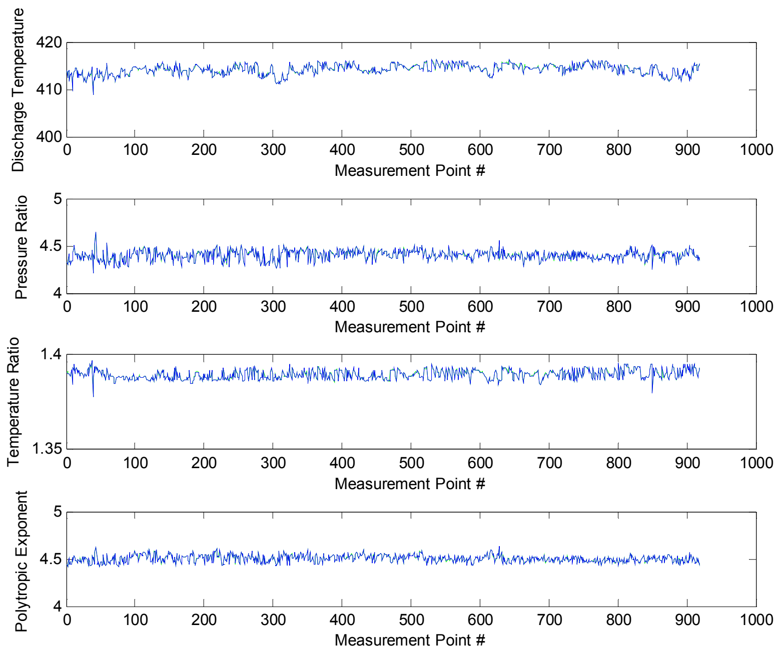


Figure 5-22: Ideal Performance Parameters Prediction 2

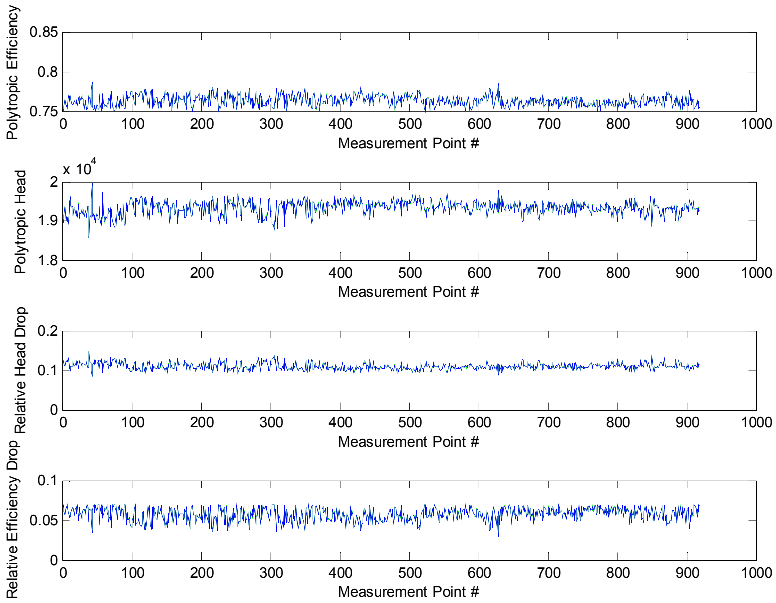


Figure 5-23: Ideal Performance Model Prediction 3

Using a different dataset, the normalised fuzzy weighted distance is calculated by fuzzifying the input data using rectangular membership functions and the cluster centres obtained from the normal operating envelope matrix, other membership functions like the Gaussian curve, triangular and sigmoid membership functions can be also be used. The idea is that using a hybrid of ECM and normalised fuzzy weighted distance measure, the ideal behaviour representing where the machine is expected to be running will be estimated and compared with where the machine is actually running through a tracking signal method. Figure 5-24 shows the relative efficiency drop actual and ideal prediction using the above method. The bottom trend in Figure 5-24 shows the residual representing the difference between the ideal predicted value and the actual value. Upper and lower limits (the black dashed lines) can be configured, when the behaviour of the equipment departs away from the normal operating envelope the model will flag that for further investigation. The tracking signal is used to monitor this process. Another example is shown in Figure 5-25 for the Polytropic exponent parameter. The orange bar in both residual plots is showing the period within which the parameters are outside the normal operating envelope.

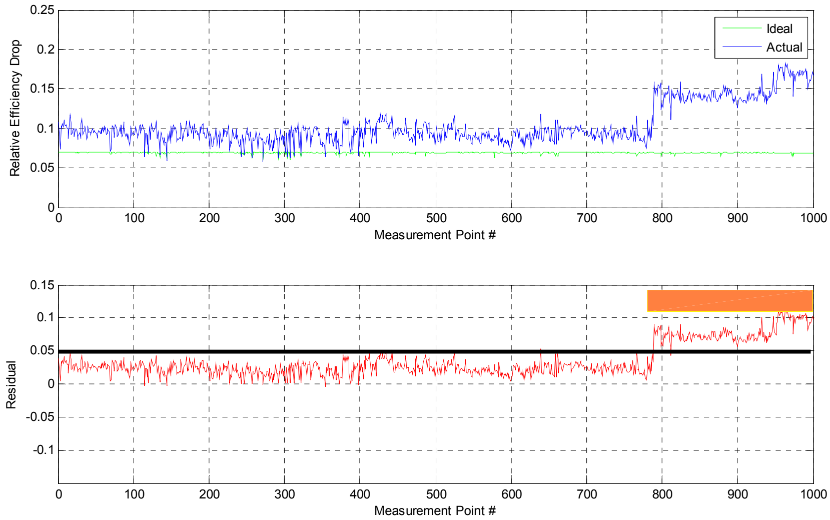


Figure 5-24: Relative Efficiency Drop Actual and Ideal Prediction

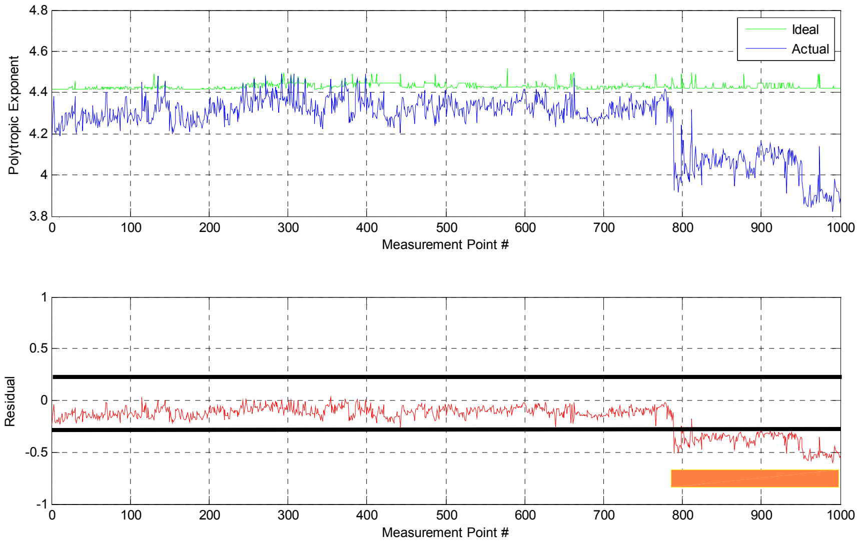


Figure 5-25: Polytropic Exponent Actual and Ideal Prediction

Model Performance Metrics

The following performance metrics were used in (Wegerich 2006) to measure the performance of the models and will be used here for comparison purposes also: Robustness, Spillover and Error. These performance metrics are described in more details in Chapter 4.2.2. Three different methods are tested: Proposed Model with rectangular membership functions, Proposed Model with exponential membership functions and the Generalized Regression Neural Network (GRNN). The following tests were conducted:

1- Test over all 919 data pairs for all three methods to predict the ideal estimate and measure the error using each test.
2- Test over 400 data pairs for all three methods to predict the ideal estimate.
3- Same as 2 however a small change (0.05) was added after normalising the data to each variable at a time and the estimates are recorded for each run. The robustness and spillover measures are then calculated.

The Polytropic exponent, relative head drop and relative efficiency drop ideal estimates and actual data are shown in figures 5-26, 5-27 and 5-28. In addition; the Error performance measure plot for all three methods is shown in figure 5-29. The first two methods i.e. the proposed model using both rectangular and exponential membership functions showed comparable error measure with better performance obtained using the exponential membership functions. The worst performance is obtained using GRNN.

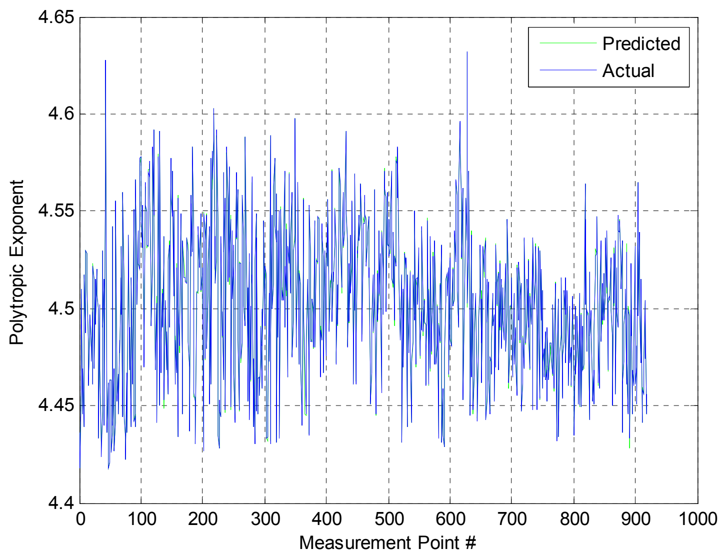


Figure 5-26 Polytropic Efficiency Actual and Predicted Trends

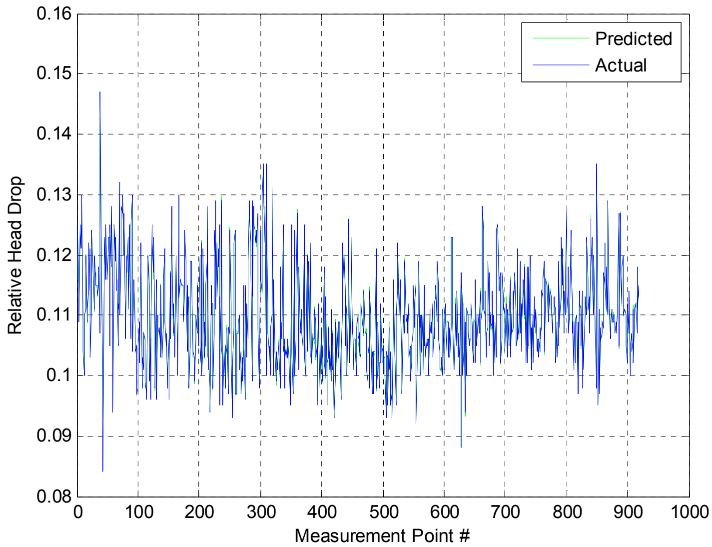


Figure 5-27 Relative Head Drop Predicted and Actual Trends

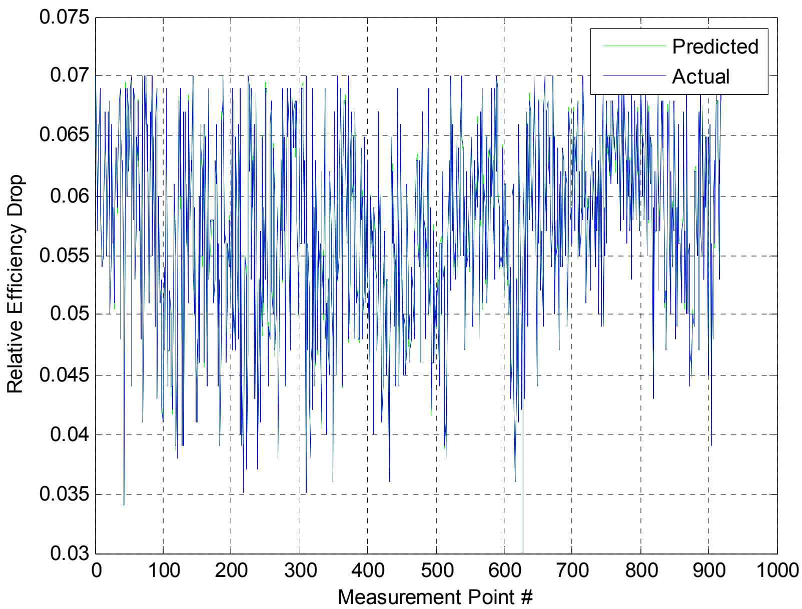


Figure 5-28 Relative Efficiency Drop Predicted and Actual Trends

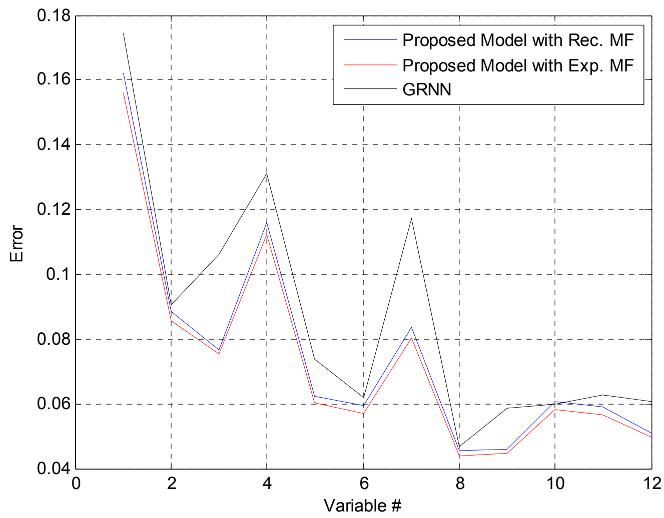


Figure 5-29 Error Performance Measure Comparison

Figures 5-30 and 5-31 show the average robustness and spillover for each of the three methods respectively. The spillover results for both method 1 and 2 showed comparable values however at least 48% improvement when compared with GRNN. On the other hand, the robustness values showed significantly better performance from the proposed model using exponential membership functions when compared with the proposed model using rectangular membership functions and GRNN.

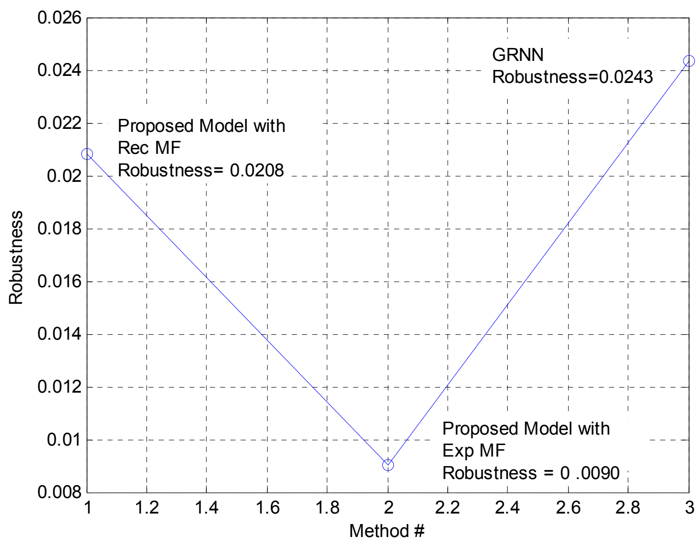


Figure 5-30 Average Robustness Measure Comparison

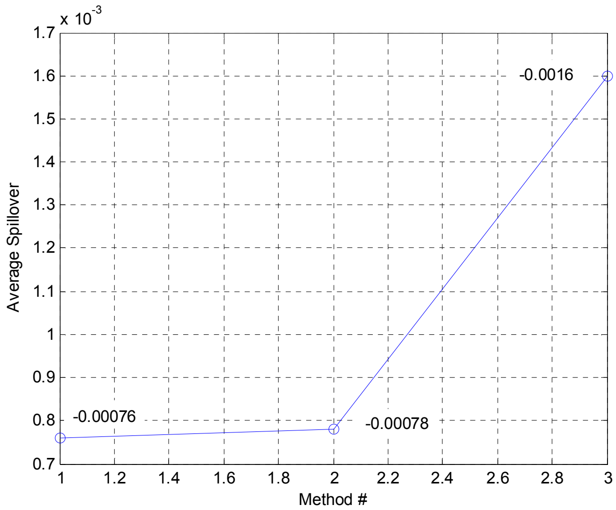


Figure 5-31 Average Spillover Measure Comparison

Figure 5-32 shows a summary of the performance metrics results obtained for all three methods. The proposed model with exponential membership functions outperformed the other two methods.

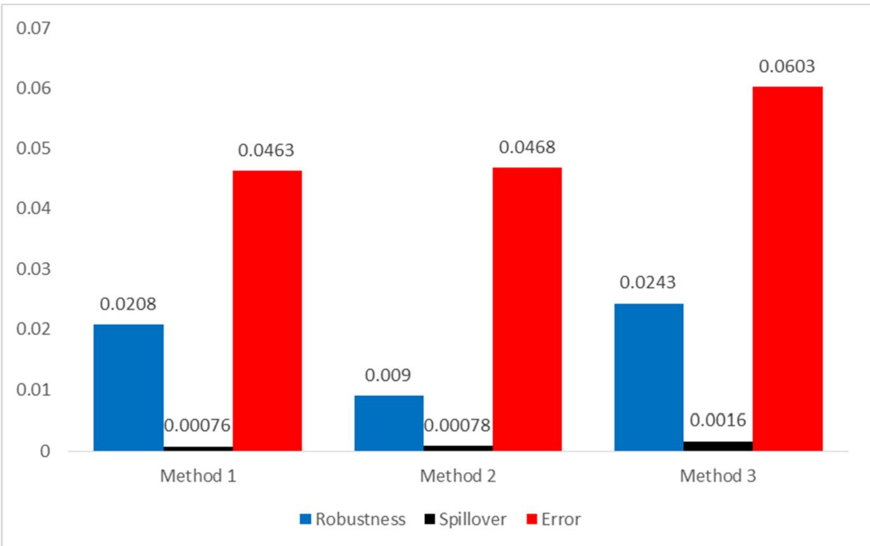


Figure 5-32 Summary Performance Metrics Results

5.5.3 Case 2: Fault Classification and Diagnostics

The proposed automated classification and diagnostics system is built based on the extended ECM algorithm for classifications (Song and Kasabov 2001a). The main idea is to train ECMs algorithm using inputs/output pairs, here the output is the known class of the data pair whether it's a fault or a no fault class. The method is described in more details in Chapter 3 with example benchmark test using the Iris dataset in Chapter 4. The ECMs algorithm was modified to allow further update to the clusters location and size during the testing stage (as required) if a data pair example arrived and is found to be outside all the available clusters i.e. doesn't belong to any of the currently available clusters. This modification is targeted at enhancing the accuracy of the method if not enough data are available for learning the model. The output of this model during testing is either one of the three below statuses:

1- Normal Mode. No Fault Identified.
2- Fault Mode (with a specific failure mode type).
3- New Mode (with unknown failure mode type).

The third output is a result of the machine operating in a new mode. This will be flagged to the user for further investigation. Findings of this investigation are fed back to the system by either assigning a specific failure mode type to this new mode or considering this mode as a normal one.

Due to the limited information available on this compressor which cover only one failure mode i.e. fouling. The field dataset used here to test the proposed diagnostics and classification method isn't for this compressor. Infact the only data that is available is for a different type of equipment. The field dataset used here is collected from several sources and is presented in (IEC 60599 1999) and (Duval and dePablo 2001) for Transformers faults classification detectable by dissolved gas analysis (DGA) and confirmed through visual inspection. The concentration in ppm of a number of gases is presented including: H_2, CH_4, C_2H_4, C_2H_6, CO and CO_2 for the below coded faults:

1. PD=partial discharges
2. D1 = discharges of low energy
3. D2 = discharges of high energy
4. T1/T2 = thermal faults of temperature <300 ºC and thermal faults of temperature 300 ºC<T< 700 ºC
5. T3 = thermal faults of temperature > 700 ºC.
6. NF= No Fault

The following inputs are used in the model testing, this will allow a comparison to be made with other researchers' results (IEC Publication 599 1978), (Huang and Huang 2002), (Ganyun et al. 2005) and (Dong et al. 2004) using the same dataset: Inputs: C_2H_2/ C_2H_4, CH_4/H_2, C_2H_4/C_2H_6.

The output is either one of the following: 1 (PD), 2 (D1), 3 (D2), 4 (T1/T2), 5 (T3) and 0 (No Fault). The distribution of the data is shown in Table 5-3. The dataset used in the form of 3 inputs-1 output are listed in Table A.1 (Appendix 1).

Table 5-3 Fault Codes and Their Distribution within the Dataset

Fault Code	Sample #
PD	1 to 6
D1	7 to 29
D2	30 to 74
T1/T2	75 to 89
T3	90 to 105
NF	106 to 134

Both training and testing datasets contains 134 data pairs (inputs/output) randomly selected from the above table. The testing showed 7 misclassified cases out of 134 case (94.78%). The lowest accuracy obtained with the PD fault class 1 since only 6 cases were included during the learning stage. Results are shown in Figures 5-33 & 5-34. Only one case was diagnosed as a no fault case whereas it was infact a problem (Partial Discharge Case). In addition to only one case that was diagnosed as a problem (T1/T2 thermal fault) when it is actually a no fault case.

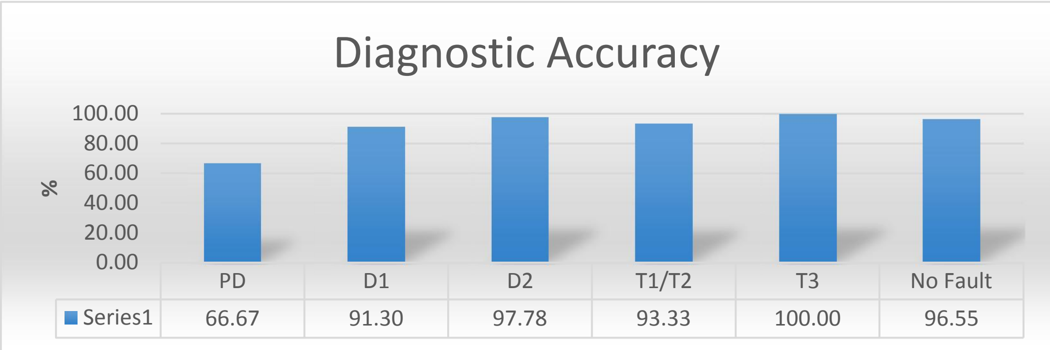


Figure 5-33 Diagnostics Accuracy Measure for the Transformers Faults

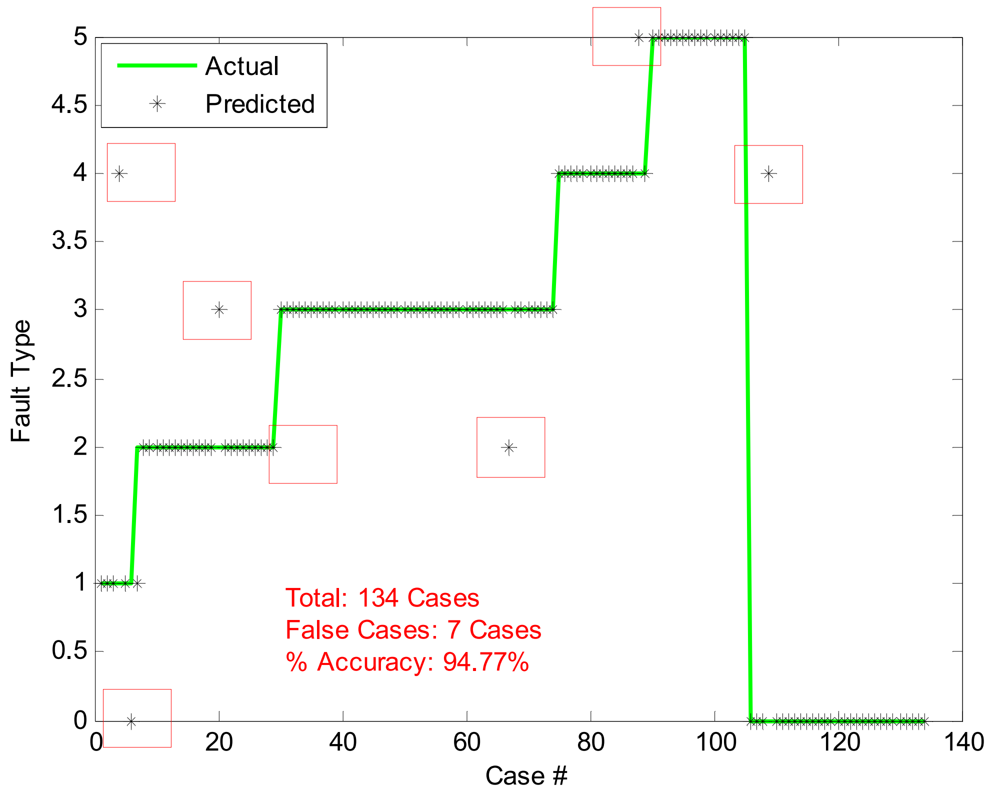


Figure 5-34 Transformers Testing Actual and Predicted Outputs

Without the modifications included in terms of updating the model during the testing stage when a new data pair doesn't lie within any of the available classes, the same test was conducted and 14 misclassified cases were obtained (89.9%) as compared to 7 misclassified cases with the modified algorithm. A comparison was made with other methods used in the literature to test the same data showed very close performance from the Evolving Wavelet Network (EWN) method (Huang and Huang 2002), all the other methods showed lower performance between 1.5% and 12.5%.

Table 5-4 Faults Classification Accuracy Comparison

Method	**Accuracy%**
IEC/IEEE std (IEC Publication 599 1978)	82.22%
ANNs (Ganyun et al. 2005)	89.05%
SVM-KNN (Dong et al. 2004)	93.10%
EWNs (Huang and Huang 2002)	94.60%
ECMs (without modifications)	89.90%
ECMsm (proposed model)	94.78%

5.5.4 Case 3: Short Term Fault Progression Prediction

Once the behaviour of the machine represented by the condition monitoring parameters (health indicators) started to drift outside the NOE, it is important to predict the future behaviour of the machine and the progression of the fault to enable planning the maintenance task required in terms of time, scope of work, resources and materials, however this step can also be done even if no fault is identified at present. This prediction is normally done using historical and current values of the condition monitoring parameters (health indicators) to predict future values. Three main steps are included within the short term prediction proposed model:

1- MDENFIS Model Training: This will be in the form of selecting initial number of data pairs from the training dataset to obtain initial number of clusters and radii. This will then be followed by online training and recursive least square update for the consequent parameters. Once the online training is finished. Modified Particle Swarm Optimisation is then used to optimise the location of the cluster centres and width of the membership functions.

2- Next Step Ahead Prediction: Online testing for next step ahead prediction is conducted in this step.

3- Online MDENFIS Updates: checking for new clusters not captured during the learning stage is conducting if a new cluster is added this will be updated by adding a new rule to the set of rules. New data arrives and testing using all new rules. This is an ongoing process. No restriction is made here for adding new clusters as the main objective of this step is to predict next step ahead and detect any transient changes in the process.

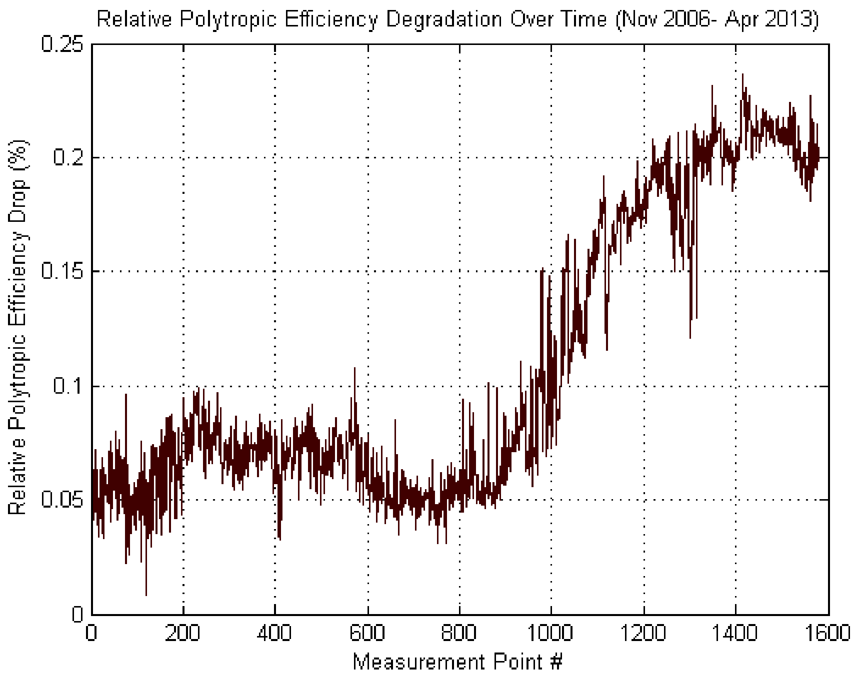


Figure 5-35 Relative Efficiency Drop Over Time

The relative efficiency drop trend over the period 2006-2013 is shown in Figure 5-35. This represent the deterioration trend since the data is made available in the data historian up until the planned bundle change in which a heavily fouled compressor was discovered. A picture of the fouled compressor is shown in figure 5-10; this parameter is considered as the main health indicator for the fouling deterioration progression rate and will be used here as an output to the model.

The inputs to the model were select from Table 6-6 based on their correlation with the output relative efficiency drop. Any input with correlation value less than 0.75 is eliminated from the test. As can be seen below, the relative head, suction pressure, discharge temperature, temperature ratio, head, pressure ratio, Polytropic efficiency and Polytropic exponent have higher than 0.75 correlation and as such they are included in the test. The rest of the parameters aren't included. The pressure and temperature ratios in addition to the Polytropic exponent trends are show in Figure 5-36.

Table 5-5 Inputs Correlation Results with the Relative Efficiency Drop

	Relative Efficiency		Relative Efficiency
Relative Efficiency	1	**Head**	-0.9302
Relative Head	0.9152	**PR**	-0.9526
Psuc	0.8683	**Polytropic Exponent**	-0.9729
Tdis	0.8234	**Polytropic Efficiency**	-0.9768
TR	0.7516		
Cpav	0.7304	**Pdis**	-0.6315
Zav	0.5272	**kav**	-0.6884
Tsuc	0.3886	**Flow**	-0.1247

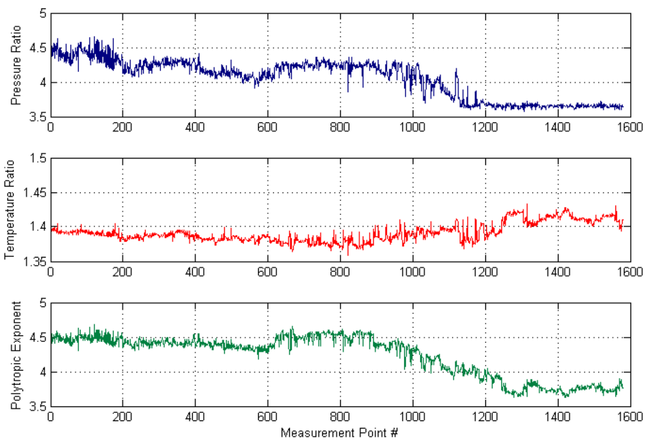


Figure 5-36: Pressure Ratio, Temperature Ratio and Polytropic Exponent Trends

Table 5-6 shows the results of the different simulations.

Table 5-6: Training and Testing NDEI and RMSE Error Measures

Results	**Training**		**Testing**	
	NDEI	**RMSE**	**NDEI**	**RMSE**
Test 1: MDENFIS w/o MPSO	0.0978	0.0016	0.1481	0.0046
Test 2: MDENFIS with MPSO	0.0362	0.00058	0.1109	0.0037
Test 3: MDENFIS with MPSO and Online Update	0.0362	0.00058	0.1035	0.0034

The database was splitted into training (1000 data pairs) and test (580 data pairs). At the beginning of the test 108 data pairs were selected to get the initial number of clusters. 4 clusters were created from this run with initial NDEI and RMSE of 0.0031 and 0.000041 respectively. The remaining 892 data pairs were used during the online training. Additional 13 clusters were added. With the first test following the online training, the test dataset was used to predict the next step ahead for the relative efficiency drop. In the second test before starting the online testing the model was taken offline for MPSO in order to optimise the cluster centres and membership functions widths before starting the online testing. The third test is the same as the second test however the online update feature was enabled while testing is ongoing. 14 additional clusters were added during the online testing. The training NDEI and RMSE are the same for tests 2 & 3 and are better 2.7 times the training NDEI and RMSE for test 1, this is a result of adding the MPSO and enabling the online updating feature. Test 3 showed better performance during the testing stage as a result of the additional clusters updated during the online testing.

100 generations (iterations) were used during the MPSO to optimise the clusters centres and membership functions widths. 5 particles were chosen for this purpose. Enhanced accuracy can be obtained by optimising the MPSO parameters using for example more iterations. Figure 5-37 shows the training

NDEI changes over all 100 iterations. The accuracy was improved by 50% over the first 2 iterations, then continued to improve rather slowly after that.

The three tests results overlaid in the same figure as the original testing data are shown in Figure 5-38. A zoom between 450 and 500 data pairs is shown in figure 5-39 and is clearly indicating the improvement in the prediction accuracy between test 1 and test 2&3.

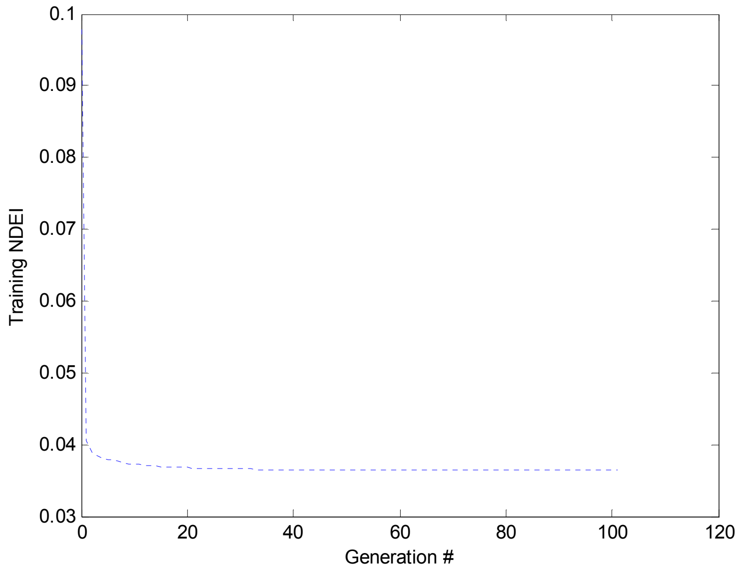


Figure 5-37 Training NDEI Changes over 100 Generations

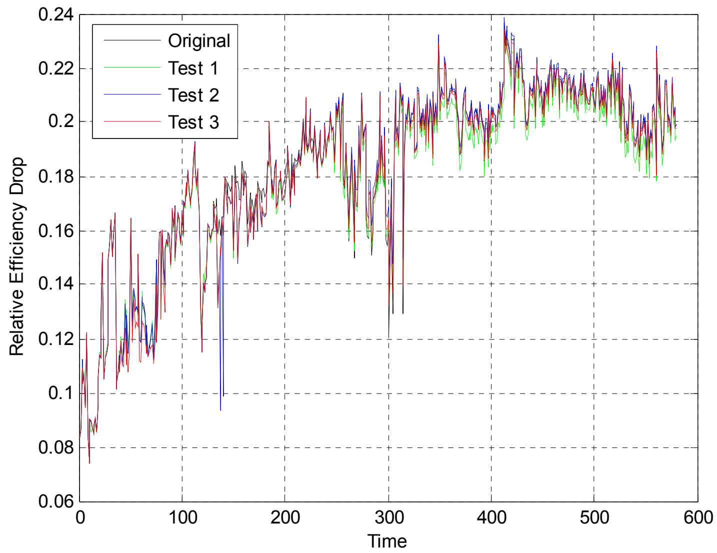


Figure 5-38 One Step Ahead Prediction Results Using Test 1,2 and 3

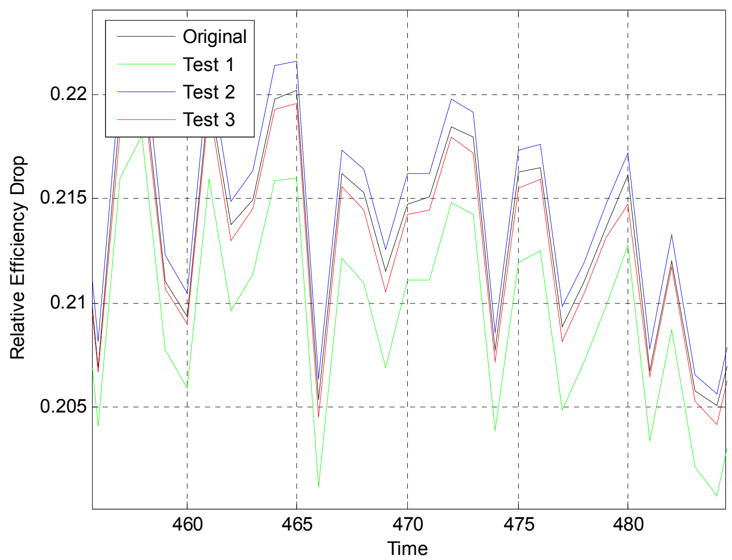


Figure 5-39 Zoomed Results Within the following (455 and 485 Steps)

If future inputs data pairs are available at present, multiple step ahead prediction (long term prediction) can be obtained with high accuracy, infact as the testing inputs are available over the remaining useful life, the above simulation can be considered remaining useful life prediction (prognostics). However since normally this data isn't available only single step ahead prediction can be made unless the inputs are predicted and then input back into the model for long term prediction purposed. Next section will describe a method for short term and long term prediction assuming inputs aren't available and will need to be predicted before the output (relative efficiency drop) can be predicted.

5.5.5 Case 4: Long Term Fault Progression Prediction

The MDENFIS trained previously is used with the short term and long term predicted inputs to predict the output short term and long term, respectively. The tracking signal will still be used during this stage for the other health indicators to detect any early signs of drifting from the normal behaviour.

The first step in the process is to apply the empirical mode decomposition method to extract the main IMFs (Intrinsic Mode Functions) for each of the inputs and output. This is then followed by short term and long term prediction for the inputs using ARIMA. MDENFIS is used to train the model using the available training inputs/output pairs. Then the predicted inputs are used to predict the output (relative efficiency drop).

Figure 5-40 shows an example of the results of the empirical mode decomposition. All the IMFs extracted from the original signal (output) are shown.

Only the 4 major IMFs are used and their sum overlaid on the same figure as the original output is shown in Figure 5-41.

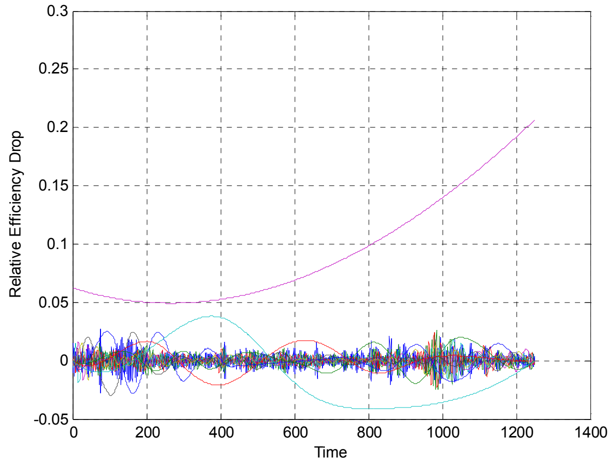


Figure 5-40 Empirical Mode Decomposition Results for the Relative Efficiency Drop Trend

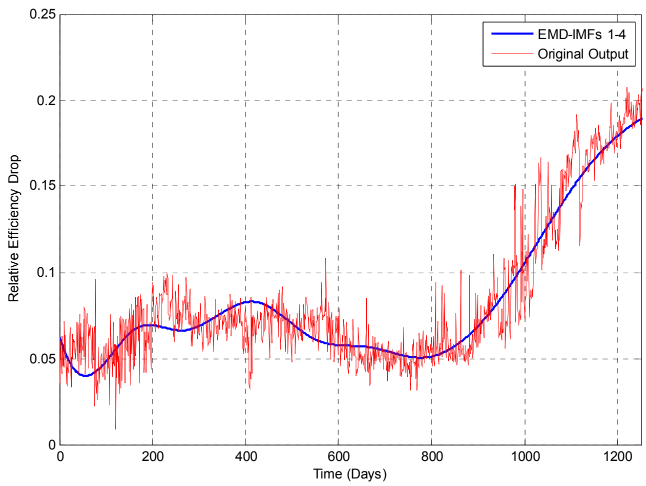


Figure 5-41 EMD Results for the Sum of the first 4 IMFs Including the Residual

The remaining part of the signal considered as noise is shown in Figure 5-42 and equates to about 0.014 of RMSE. This is the same as 0.014 error in the relative efficiency drop which is considered acceptable.

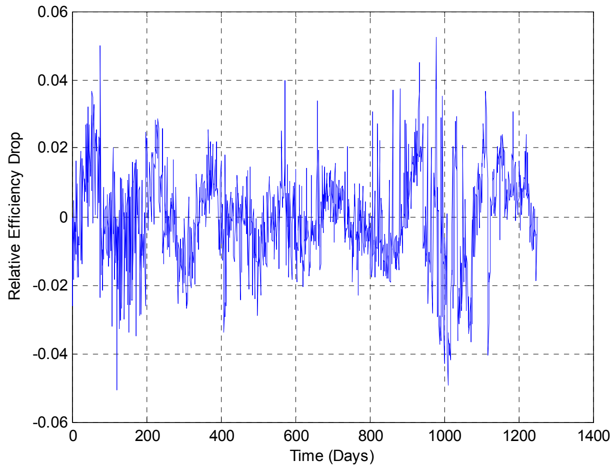


Figure 5-42 Remaining Noise of the Relative Efficiency Drop

The major IMFs were extracted for all inputs and are summed together for each input, then an automatic arima modeller was used to select the best arima model lags based on BIC (Bayesian information criterion). The following ARIMA models were selected for the inputs 1-8 respectively; ARIMA(2,1,1), ARIMA(1,2,1), ARIMA(1,1,1), ARIMA(1,2,1), ARIMA(1,2,1), ARIMA(1,2,1), ARIMA(1,2,1) and ARIMA(1,1,1). 250 steps ahead were predicted using ARIMA for each one of the inputs. Then, MDENFIS was used to predict the output relative efficiency drop 250 steps ahead. Figure 5-43 shows the results of this test. The calculated RMSE and NDEI are 0.0063 and 0.2156 respectively.

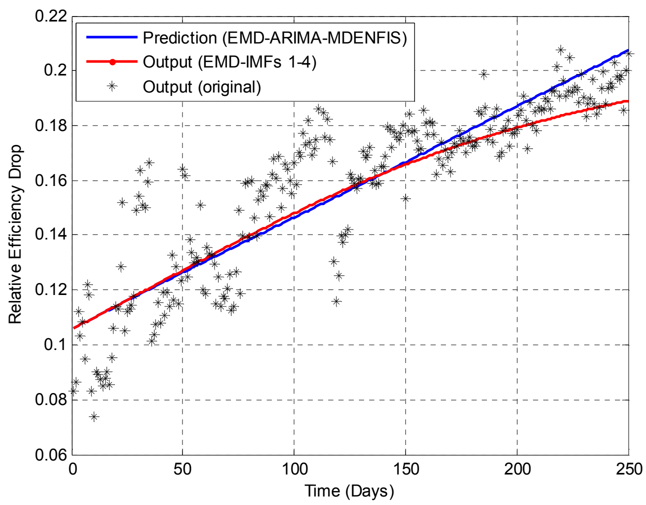


Figure 5-43 250 Step Ahead Prediction Using EMD-ARIMA-MDENFIS

5.5.6 Summary

This chapter covered the implementation of the proposed integrated condition monitoring and prognostics framework on field datasets. The implementation was successful by means of comparison with methods developed by other researchers and comparing the performance metrics like NDEI, RMSE, MAE, Robustness, Spillover, Error, etc. Four case studies are presented to test every part of the proposed model starting from the normal operating envelope going through the initiation of fault and the fault classification and diagnostics and finally the short term and long term prediction.

5.6 HYPOTHESIS TESTING OUTCOMES

In light of the outcomes from testing the proposed model on benchmark datasets (Chapter 4) and actual field data (Chapter 5) and the comparison with other researchers' results, the hypothesis presented in Chapter 1 and Section 4 have been met as follow:

1. Significant improvement in the prediction accuracy and processing time as a result of the proposed modifications to the original model (>50%). This hypothesis is validated and compared using Benchmark and field datasets.

2. The model is able to detect sudden changes in the process and add new rules consequently.

3. The use of MPSO has improved the prediction accuracy of the proposed model.

4. The proposed model is able to address the lack of future input data to the model by predicting this data and consequently inserting them back into the model.

5. The proposed modifications to the original classification algorithm MCMc improved the classification accuracy of the model.

6 Conclusions and Recommendations

This chapter first summarises the conclusions in light of the benchmarks and field datasets case studies results from chapters 4 and 5. Main contributions of this work are listed in section 6.2. The chapter concludes with future work suggestions.

6.1 Conclusions

The increasing need of manufacturers for highly reliable operation of their machinery and minimum production deferrals due to unplanned downtimes has put alot of pressure on the maintenance managers and researchers to continuously develop new techniques and strategies to meet such objectives. This is evident by the recent efforts in moving from the traditional corrective maintenance approach to more comprehensive approaches like preventive (timely based) and predictive (condition based) maintenance protocols. The literature review conducted between 1966 and 2013 has shown an exponential increase in publications in the field of condition monitoring of rotating equipment, Figure 6-1 recognising the importance of this area as part of any maintenance regime.

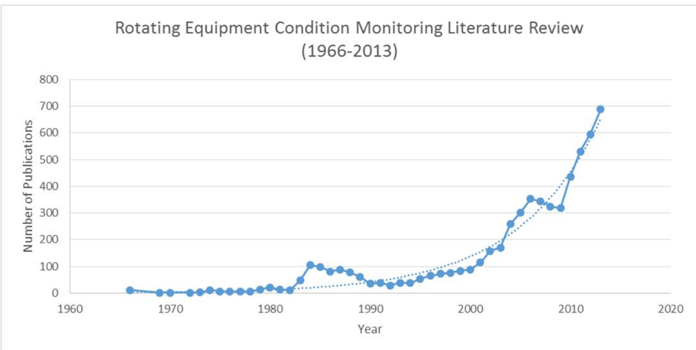


Figure 6-1 Rotating Equipment Condition Monitoring Literature Review

Recent trends are now encompassing capabilities for ascertaining the remaining useful operational life (prognostics) giving maintenance managers even more information for scheduling the appropriate maintenance actions at least cost. Different methods are used which vary in complexity and amount of data required. Models utilizing reliability data, CM data and/or physics based principles were developed which in some applications proven to be good approaches however each of these models has its merits and disadvantages. Similar exponential trend was noticed during the last 30 years review of publications in the field of machinery prognostics including journal articles, conference proceedings, etc., Figure 6-2.

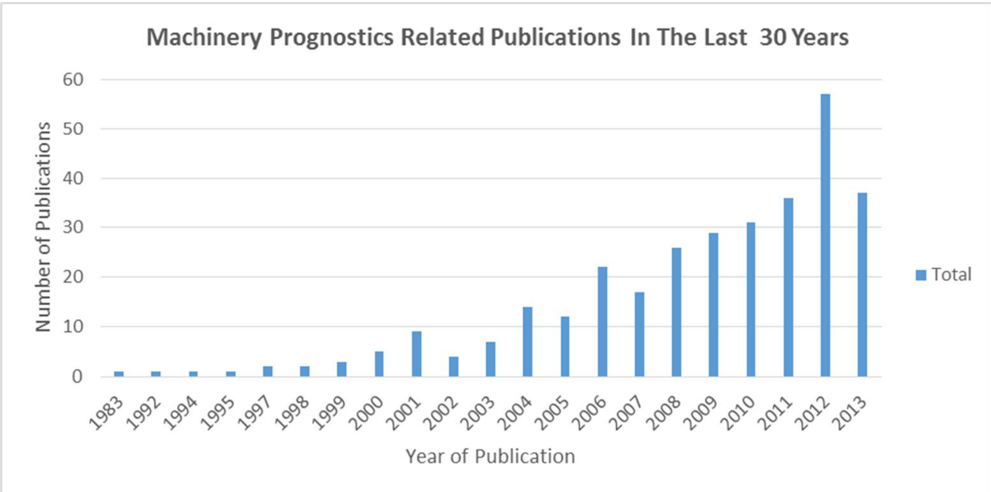


Figure 6-2 Machinery Prognostics Related Publications in the Last 30 years

The following conclusions highlight the results of this work:

- An integrated condition monitoring and prognostics framework was proposed. This framework is fully aligned with ISO 17359:2011 (BS ISO 17359:2011 2011) and is based around the I-P and P-F curve but more specifically around the life cycle of the equipment.

- 6 Main steps were discussed within this framework, with more emphasis on detecting departure from the NOE, classification and diagnostics, short term and long term predictions. The main steps are: Step 1-Data collection and validation, Step 2-data processing and feature extraction, Step 3-fault classification and diagnostics, Step 4- fault progression prediction, Step 5-maintenance action and Step 6-continuous improvement.
- An online condition monitoring and prognostics model is proposed to support the above framework, which is highly flexible to support a number of applications including prediction and classification. This model is inspired by the evolving connectionist system theory.
- A hybrid of ECM and normalised weighted fuzzy distance measure was proposed to monitor where the machine is running with respect to its NOE and detecting early stages of departure outside this zone using a tracking signal.
- A modified version of the ECM for classification algorithm is proposed for the automated fault classification and diagnostics step. The proposed change is aimed at allowing further update to the clusters (fault or no fault status) location and size during the test stage if new data arrives and is found outside any of the regions covering the existing clusters.
- A modified version of the dynamic evolving neuro-fuzzy inference system (one of the evolving connectionist systems models) is presented here called MDENFIS. This model is used for the short term and long term prediction.

- MDENFIS includes the following as part of its structure: an ECM module to grid partition the searching space, weighted recursive least square method to solve the consequent unknown parameters of the model, modified particle swarm optimisation to optimise the location of the clusters obtained by the ECM and the width of the membership functions, an online update function during the testing stage to increase the model's sensitivity of any genuine transient events. All covered in a NeuroFuzzy framework to learn the nonlinear relationship between the inputs and output through a hybrid Neural Networks and Fuzzy Logic data driven model.
- Almost all the publications in the field of time series prediction and prognostics, assumed that the future inputs to the model are available at time of testing the model by means of splitting the available dataset into two or three separate databases, one for training the model, one for validating it and potentially a third one for testing it, although majority have two different dataset for training and testing purposes. This isn't the case in reality especially when trying to implement the proposed model on a newly installed machine where no previous history is available.
- Taking into consideration the above point about the inputs future availability. There need to be an option to predict the future values of the inputs, prior to testing that on the MDENFIS model for short term and long term predictions. As such, an integrated EMD-ARIMA model was proposed to predict the future values of the inputs after which these predicted inputs will be inserted into the MDENFIS model during the testing stage.

- Testing the model on a benchmark dataset has shown improvement in the prediction accuracy when compared with other researchers. Since the model is automatically updating its structure while online on the testing stage this will allow the model to learn any events that were missed during the learning stage giving rise to a wider range of applications including machinery with limited or no historical data, or a minimum amount of training data. Further developments on the model are currently in progress.
- The proposed model was first tested on three benchmark datasets, namely: Mackey-Glass Chaotic Time Series Dataset 1, Gas Furnace Dataset 2 and Iris Flower Dataset 3. Those three datasets were used by other researchers as indicated in this work for similar purposes including prediction for the first two datasets and classification for the third one. This will enable a comparison of like to like to be made by testing the performance of the proposed model against other researchers work using one or more of the following performance metrics: MSE, RMSE, MAE, Robustness, Spillover, Error, NDEI and processing time.
- The benchmark testing for the prediction capabilities of the model has shown that the MDENFIS model improved the training prediction accuracy by 60% and the testing prediction accuracy by 50% when compared with DENFIS, with 55 rules compared to 883 rules with DENFIS, this is obviously an improvement on the processing time during the testing stage. Case Study No.3 in section 4.5 covers more details. The model has also capability of detecting sudden changes during the testing stage online by

adding additional clusters to the model which weren't there during the training stage, section 4.6.

- The benchmark testing for the proposed model classification and diagnostics capabilities using Iris flower dataset 3 has shown at least 96.66% classification accuracy while misclassifying the class of two cases out of 75 which is higher than the classification accuracy of some of the known classification algorithms as shown in section 4.7.
- The proposed model was successfully tested on field data for a fouled compressor. The model was able to detect at an early stage the deviation from the NOE of various performance parameters of the compressor and was able to predict 250 days ahead of the fouling progression rate by means of the relative efficiency drop as the health indicator (the output of interest).
- The proposed model was tested on various faults of electric transformers. The model misclassified 7 cases out of 134 case giving a prediction accuracy of 94.78% with the next better accuracy of 94.60% using evolving wavelet networks.

6.2 Contributions

The main contributions of this work are:

1. Proposing a hybrid of ECM and normalised weighted fuzzy distance for the first time to monitor where the machine is running with respect to its normal operating envelope (NOE) and then alert the user if departure outside this zone occurs by using a tracking signal. The combination of ECM and weighted fuzzy distance is in its own a new model and also the

usage of this model for predicting the normal equipment behaviour and compare that with where it's actually running is a second contribution here. The process is proposed to be monitored using a tracking signal which has the capability of identifying any departure beyond the NOE.

2. Modifying the ECM for classification algorithm to enable online updating of the clusters during the testing stage. This feature enhanced the classification accuracy of 150 cases of the Iris flower from 98.6% to 100%.
3. Enhancing the prediction accuracy of the Dynamic Evolving NeuroFuzzy Inference model by introducing the following changes to the model: using exponential membership functions instead of the originally proposed rectangular membership functions, integrating the model with a modified version of particle swarm optimisation to optimise the position of the clusters centres and width of the membership functions, and finally adding an online updating feature during the testing stage online to enhance the models capability of detecting sudden changes in the process.
4. Introducing a hybrid EMD-ARIMA model to enable predicting the future values of the inputs which in turn enables prediction of multiple steps ahead (long term prediction) of the health indicators. This modification was successfully tested on fouling progression rate and the model predicted successfully 250 days steps ahead.

6.3 Future Work

The following tasks are part of potential future work in this research:

- Testing the proposed model on other type of equipment, faults and CM data like vibration and lube oil analysis to evaluate the possibility of generalising this approach for prediction and classification of any type of equipment (rotating, reciprocating and static equipment).

- Building a Multi-Input- Multi-Output (MIMO) MDENFIS model that can predict the time to failure of machinery.
- Improving the prediction accuracy of the model by deciding which inputs to be included during the prediction online using iterative correlation or PCA models.
- Add the capability of the model to integrate condition monitoring and reliability data.
- Test/Validate the model using simulated CM data and apply it to real world case studies when there are limitations in the data.
- Integrating the proposed condition monitoring and prognostics model with a maintenance regime to enable coordination between the outcomes of this model and the maintenance work to be done but also looking at critical spares availability.

REFERENCES

Abdel-Magied, M.F., Loparo, K.A. and Lin, W. (1998), "Fault Detection and Diagnosis of Rotating Machinery", in: American Control Conference, 1998. Proceedings of the 1998, Philadelphia, PA, USA, vol.6, pp.3291-3296.

Absolute Astronomy (2009), “Exploring the universe of knowledge: Prognosis, available at: http://www.absoluteastronomy.com/topics/Prognosis# encyclopaedia

Al-Hassan, W., Fayek, M.B. and Shaheen, S.I. (2007), “Psosa: An optimized particle swarm technique for solving the urban planning problem”, In Computer Engineering and Systems, The 2006 International Conference on, pp. 401–405.

Amarnath, M., Sugumaran, V. and Kumar, H. (2013), "Exploiting sound signals for fault diagnosis of bearings using decision tree", Measurement, vol. 46, no. 3, pp. 1250-1256.

Anderson, O. 1976, Time series analysis and forecasting: the Box-Jenkins approach: London, Butterworths, pp. 182.

Artes, M. (2007), "Application of Cluster Analysis to Machine Fault Diagnosis using Vibration Signal Processing", Machine Design, pp. 377-380.

Arumugam, M.S. and Rao, M.V.C. (2006), “On the performance of the particle swarm optimization algorithm with various Inertia Weight variants for computing optimal control of a class of hybrid systems”, Discrete Dynamics in Nature and Society.

ASME PTC-10 (1997), “Performance test code on compressors and exhausters”, American Society of Mechanical Engineers, New York.

Baillie, D. C. and Mathew, J. (1996), "A COMPARISON OF AUTOREGRESSIVE MODELING TECHNIQUES FOR FAULT DIAGNOSIS OF ROLLING ELEMENT BEARINGS", Mechanical Systems and Signal Processing, vol. 10, no. 1, pp. 1-17.

Bain, A. (1873), “Mind and body. The theories of their relation”, London: Henry King

Balocchi, R., Menicucci, D., Santarcangelo, E., Sebastiani, L., Gemignani, A., Ghelarducci, B. and Varanini, M. (2004), "Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition", Chaos, Solitons & Fractals, vol. 20, no. 1, pp. 171-177

Bansal, J.C., Singh, P.K., Saraswat, M., Verma, A., Jadon, S.S. and Abraham, A. (2011), "Inertia Weight strategies in Particle Swarm Optimization," Nature and Biologically Inspired Computing (NaBIC),Third World Congress on , vol., no., pp.633-640.

Barakat, M., Druaux, F., Lefebvre, D., Khalil, M. and Mustapha, O. (2011), "Self adaptive growing neural network classifier for faults detection and diagnosis", Neurocomputing, vol. 74, no. 18, pp. 3865-3876.

Barnhart, B. L. and Eichinger, W. E. (2011), "Empirical Mode Decomposition applied to solar irradiance, global temperature, sunspot number, and CO2 concentration data", Journal of Atmospheric and Solar-Terrestrial Physics, vol. 73, no. 13, pp. 1771-1779.

Barszcz, T. and Jabłoński, A. (2001), "Aspects of automatization of wind farm monitoring on the example of a diagnostic center", Proceedings of 8th International Workshop on Structural Health Monitoring (IWSH), vol. 2, pp. 2603–2610.

Batko, W. (1984), "Prediction method in technical diagnostics", SC. Dr. Thesis, Cracov Mining Academy.

Baydar, N. , Chen, Q. , Ball, A. and Kruger, U. (2001), "Detection of incipient tooth defect in helical gears using multivariate statistics", Mechanical systems and signal processing, vol.15, pp.303–321.

Benedict, M., Webb, G.B.; Rubin, L. C. (1940), "An Empirical Equation for Thermodynamic Properties of Light Hydrocarbons and Their Mixtures: I. Methane, Ethane, Propane, and n-Butane", Journal of Chemical Physics, vol.8, no.4, pp.334–345

Berry, J. (1992); "Proven Method for Specifying Both Six Spectral Alarm Bands as Well as Narrowband Alarm Envelopes Using Today's Predictive Maintenance Software System 1 (2nd Edition) ", Predictive Maintenance and Vibration Signature Analysis II; technical Associates of Charlotte, NC.

Bin, G. F., Gao, J. J., Li, X. J. and Dhillon, B. S. (2012), "Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network", Mechanical Systems and Signal Processing, vol. 27, no. 0, pp. 696-711.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), Time Series Analysis: Forecasting and Control, Third Edition, Englewood Cliffs, NJ: Prentice Hall, pp.197–199.

Box, G.E.P, and Jenkins , G.M., (1976), Time Series Analysis: Forecasting and Control, (2nd ed.), Holden-Day, San Francisco.

Brown , R. J.C. , Goddard, S.L. and Brown, A.S.(2010), "Using principal component analysis to detect outliers in ambient air monitoring studies", International Journal of Environmental Analytical Chemistry, vol.90, no.10, pp.761-772.

Brown, R. G. (1959), Smoothing, Statistical Forecasting For Inventory Control. McGraw-Hill, New York.

Brown, R.G. and Meyer,R.F. (1961), "The Fundamental Theorem of Exponential Smoothing", Operations Research. vol.9, no.5, pp. 673-685.

BS 5760-5:1991 (1991), Reliability of systems, equipment and components. Guide to failure modes, effect and criticality analysis (FMEA and FMECA).

BS ISO 13372:2012 (2012), “Condition monitoring and diagnostics of machines — Vocabulary”.

BS ISO 17359:2011 (2011), “Condition monitoring and diagnostics of machines — General guidelines”.

Cadini, F. Zio, E. and Avram, D. (2009), "Monte Carlo-based filtering for fatigue crack growth estimation", Probabilistic Engineering Mechanics, vol.24, no.3, pp.367-373.

Caesarendra, W., Niu, G. and Yang B.-S. (2010),"Machine condition prognosis based on sequential Monte Carlo method", Expert Systems with Applications, vol.37, no.3, pp. 2412-2420.

Cao, L. (1997), "Practical method for determining the minimum embedding dimension of a scalar time series", Physica D: Nonlinear Phenomena, vol.110, no.1-2, pp. 43-50.

Cesar, A., Hernandez, S., Luis, F., Pedraza, M., Octavio, J. and Salcedo, P. (2009), “Comparative Analysis of Time Series Techniques ARIMA and ANFIS to Forecast Wimax Traffic”, The Online Journal on Electronics and Electrical Engineering,vol.2, no.2.

Chang, R., Wu, W., Moon, W. and Chen, D. (2003), "Improvement in Breast Tumor Discrimination by Support Vector Machines and Speckle-Emphasis Texture Analysis", Ultrasound in Med. & Biol., vol. 29, no. 5, pp. 679-686.

Chapelle, O. and Vapnik, V. (2001), "Choosing Multiple Parameters for Support Vector Machines", Machine Learning.

Chen, C. and Chen, J. (2011), "Rotor fault diagnosis system based on sGA-based individual neural networks", Expert Systems with Applications, vol. 38, no. 9, pp. 10822-10830.

Chen, G., Huang, X., Jia, J. and Min, Z. (2006), “Natural exponential Inertia Weight strategy in particle swarm optimization”, In Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on, vol. 1, pp. 3672–3675.

Chen,S.M. and Fang, Y.D. (2005), “A new approach for handling the iris data classification problem”. International Journal of applied Science and Engineering, vol.1,pp. 37-49.

Cheng, J., Yu, D., and Yang Y. (2006), "A fault diagnosis approach for roller bearings based on EMD method and AR model", Mechanical Systems and Signal Processing, vol.20, pp.350-362.

Cheung, C., Chui, C., and Chan, A. (1993), "Real-time detection of transient signals using Spline-Wavelets", Proceedings-ICASSP, IEEE Int. Conference on Acoustics, Speech and Signal Processing, vol. 4, pp. 361-364.

Chew, E. Y. and Wilsky, A. S. (1984), "Analytic Redundancy and the Design of the Robust Failure Detection System", IEEE Trans. Auto. Contr., vol. 29, pp. 603-614.

Chhikara, J. and Singh, J. (2012), "Noise cancellation using adaptive algorithms", International Journal of Modern Engineering Research (IJMER), vol.2, no.3, pp.792-795.

Choi, G.H. and Choi, G.S.(1996), "Application of Minimum Cross Entropy to Model-Based Monitoring in Diamond Turning", Mechanical Systems and Signal Processing, vol.10, no.5, pp.615-631.

Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots". Journal of the American Statistical Association vol. 74, no. 368, pp.829–836.

Coats, M. D., Sawalhi, N. and Randall, R.B. (2009), "Extraction of tacho information from a vibration signal for improved synchronous averaging", Proceedings of ACOUSTICS.

Coker, A. K. (2010). Ludwig's Applied Process Design for Chemical and Petrochemical Plants, Volume 2 (4th Edition). Elsevier.

Deng, L. and Zhao, R. (2013), "A vibration analysis method based on hybrid techniques and its application to rotating machinery", Measurement, vol. 46, no. 9, pp. 3671-3682.

Dong, M., Xu, D. K., Li, M. H. and Yan, Z. (2004), "Fault diagnosis model for power transformer based on statistical learning theory and dissolved gas analysis", Electrical Insulation, 2004. Conference Record of the 2004 IEEE International Symposium on, pp. 85.

Dovžan, D. and Škrjanc, I. (2011), "Recursive fuzzy c-means clustering for recursive fuzzy identification of time-varying processes", ISA transactions, vol. 50, no. 2, pp. 159-169.

Drury, J. and Cockrell, T., "A24 Antifouling Corrosion Resistant Coating", Sulzer Turbo Services.

Dunia, R., Qin, S.J. , Edgar, T.F. and McAvoy, T.J. (1996), "Identification of faulty sensors using principal component analysis", AIChE Journal,vol.42, no.10, pp. 2797 -2812.

Duval, M. and dePablo, A. (2001), “Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases,” IEEE Electrical Insulation Magazine,vol. 17, no. 2, pp.31-41.

Dybała, J. and Zimroz, R. (2014), "Rolling bearing diagnosing method based on Empirical Mode Decomposition of machine vibration signal", Applied Acoustics, vol. 77, no. 0, pp. 195-203.

Eberhart, R.C. and Shi, Y. (2002), “Tracking and optimizing dynamic systems with particle swarms”, In Evolutionary Computation, 2001. Proceedings of the 2001 Congress on, vol. 1, pp. 94–100. IEEE.

Elaine, Y.C., Wai, W.X., Michael, K.N. and Joshua, Z.H. (2004), " An optimization algorithm for clustering using weighted dissimilarity measures", Pattern Recognition, vol. 37, pp. 943–952.

Entek, "Monitor Variable Speed Machines using Narrowband Alarms", Entek Application Notes P.M. 101; Published by Entek Scientific Corportation; Cincinnati, OH.

Fahimifard, S.M., Homayounifar, M., Sabouhi, A.R. Moghaddamnia, M. (2009), “Comparison of ANFIS, ANN, GARCH and ARIMA Techniques to Exchange Rate Forecasting”, Journal of Applied Sciences, vol.9, no. 20, pp.3641-3651.

Feng, Y., Teng, G.F., Wang, A.X. and Yao, Y.M. (2008), “Chaotic Inertia Weight in Particle Swarm Optimization”, Innovative Computing, Information and Control, 2007. ICICIC’07. Second International Conference on, pp. 475. IEEE.

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems". Annals of Eugenics, vol. 7, no. 2, pp.179–188.

Forrester, B. D. (1989), "Analysis of gear vibration in the time-frequency domain", Proceedings of the 44th Meeting of the Mechanical Failures Prevention Group of the vibration Institute

Forrester, B. D. (1989), "Time frequency analysis of helicopter transmission vibration", Aeronautical Research laboratory, AR Propulsion, Melbourne, Technical Report 180.

Forrester, B. D. (1989), "Use of the Wigner Ville distribution in helicopter transmission fault detection", Proceedings of the Australian Symposium on Signal Processing and Applications.

Fung, W.K. (1999), "Outlier diagnostics in several multivariate samples", the statistician. Technometrics, vol.48, no.1, pp. 73-84.

Ganyun ,L.V., Cheng, H., Zhai, H. and Dong, L. (2005) “ Electric Power Systems Research 75,” Elsevier B.V. All rights reserved.

Gao, R.X. and Yan, R. (2006), "Non-stationary signal processing for bearing health monitoring", International J. Manufacturing Research, vol. 1, no. 1, pp.18–40.

Gao, Y., An, X. and Lie, J. (2008), “A Particle Swarm Optimization Algorithm with Logarithm Decreasing Inertia Weight and Chaos Mutation”, In Computational Intelligence and Security, 2008. CIS’08. International Conference on, vol.1, pp. 61–65.

Geitner, F. K. and Bloch, H. P. (2012), Machinery Failure Analysis and Troubleshooting (Fourth Edition), Butterworth-Heinemann, Oxford, pp. 295-390.

Geramifard, O., Xu, J. and Kumar Panda, S. (2013), "Fault detection and diagnosis in synchronous motors using hidden Markov model-based semi-nonparametric approach", Engineering Applications of Artificial Intelligence, vol. 26, no. 8, pp. 1919-1929.

Ghate, V. N. and Dudul, S. V. (2010), "Optimal MLP neural network classifier for fault detection of three phase induction motor", Expert Systems with Applications, vol. 37, no. 4, pp. 3468-3481.

Gresh, M. T. (2001). Compressor performance: aerodynamics for the user. Boston, Butterworth-Heinemann.

Gribok, A.V., Hines, J.W., and Uhrig, R.E. (2000), "Use of Kernel Based Techniques for Sensor Validation in Nuclear Power Plants", The 3rd American Neuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human-Machine Interface Technologies.

Grubbs, F. E. (1950), “Sample criteria for testing outlying observations”, The Annals of Mathematical Statistics, vol. 21, no. 1, pp.27-58.

Gryllias, K. C. and Antoniadis, I. A. (2012), "A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments", Engineering Applications of Artificial Intelligence, vol. 25, no. 2, pp. 326-344.

Hanlon, P. C. (2001), “Compressor Handbook,” McGraw-Hill.

Han-Oh, S., George, R. and Hales, R. (2012), "Analysis of Tumor-Motion Surrogate Signals Under Different Coaching Conditions Using Empirical Mode Decomposition and Hilbert-Huang Transformation", International Journal of Radiation Oncology*Biology*Physics, vol. 84, no. 3, pp. S737.

He, F.J. and Shi, W.G. (2002), "WPT-SVMs based approach fault detection of valves in reciprocating pumps", Proceedings of American control Conference.

Heng, A., Zhang, S., Tan, A. C. C. and Mathew, J. (2009), "Rotating machinery prognostics: State of the art, challenges and opportunities", Mechanical Systems and Signal Processing, vol. 23, no. 3, pp. 724-739

Herzog, J., Wegerich, S., Gross , K. C., and Bockhorst, F. K. (1998), "MSET Modeling of Crystal River-3 Venturi Flow Meters", Proceedings of ASME/JSME/SFEN 6th International Conference on Nuclear Engineering, San Diego, CA, USA.

Hong, L. (2011), "Decomposition and Forecast for Financial Time Series with High-frequency Based on Empirical Mode Decomposition", Energy Procedia, vol. 5, no. 0, pp. 1333-1340.

Huang N.E., Chern, C.C., Huang, K., Salvino, L.W/, Long, S.R. and Fan, K.L. (2001), “A new spectral representation of earthquake data: Hilbert spectral analysis of station TCU129”, Chi-Chi, Taiwan, 21 September 1999. Bull Seism Soc Amer 2001, vol. 91, no.5, pp.1310-1338.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q. , Yen, N.-C., Tung, C. C. and Liu, H. H. (1998), The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. London, Ser. A, 454, pp.903–993.

Huang, N.E., Shen, Z. and Long, S.R. (1999), "A new view of nonlinear water waves: the Hilbert spectrum", Annual Reviews of Fluid Mechanics, vol.3, pp. 417–457.

Huang, Y.-C. and Huang, C.-M. (2002), "Evolving wavelet networks for power transformer condition monitoring", Power Delivery, IEEE Transactions on, vol. 17, no. 2, pp. 412-416.

Hwang ,Y.-C. and Song, Q. (2008), “Dynamic neural fuzzy inference system”, Proceedings of the 15th international conference on Advances in neuro-information processing - Volume Part I, Nikola Kasabov, and George Coghill (Eds.), Vol. Part I. Springer-Verlag, Berlin, Heidelberg, pp.1245-1250.

IEC 60599 (1999), “Mineral oil-impregnated electrical equipment in service--Guide to the interpretation of dissolved and free gases analysis,” IEC publication.

IEC 60812 (2006), “Analysis techniques for system reliability — Procedure for failure mode and effects analysis (FMEA)”.

IEC Publication 599 (1978), “Interpretation of the Analysis of the Gases in Transformers and Other Oil - Filled Electrical Equipment in Service”.

ISO 13381-1:2004 (2004), Condition monitoring and diagnostics of machines – Prognostics - Part 1: General Guidelines.

ISO 14224 (2006), Petroleum, petrochemical and natural gas industries - Collection and exchange of reliability and maintenance data for equipment.

Jablonski, A. and Barszcz, T. (2013), "Validation of vibration measurements for heavy duty machinery diagnostics", Mechanical Systems and Signal Processing, vol. 38. pp. 248-263.

James, W. (1890), Principles of Psychology, Henry Holt, New York.

Jang, J.-S. R. (1992), “Neuro-Fuzzy Modeling: Architectures, Analyses, and Applications”, Ph.D. Dissertation, EECS Department, Univ. of California at Berkeley.

Jang, J.-S. R. (1993). ANFIS: Adaptive network-based fuzzy inference systems. IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, pp. 665–685.

Jang, J.-S. R., Sun, C.-T. and Mizutani, E. (1997), “ Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence”, Prentice Hall.

Jardine, A. K. S., Lin, D. and Banjevic, D. (2006), "A review on machinery diagnostics and prognostics implementing condition-based maintenance", Mechanical Systems and Signal Processing, vol. 20, no. 7, pp. 1483-1510.

Jegadeeshwaran, R. and Sugumaran, V. "Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines", Mechanical Systems and Signal Processing, , no. 0.

Jia, S., Howard, I. and Wang, J. (2003), "The Dynamic Modeling of Multiple Pairs of Spur Gears in Mesh, Including Friction and Geometrical Errors", International Journal of Rotating Machinery, vol.9, pp.437-442.

Jiang, L., Xuan, J. and Shi, T. (2013), "Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis", Mechanical Systems and Signal Processing, vol. 41, no. 1–2, pp. 113-126.

Jin, X. and Chow, T. W. S. (2013), "Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis–Taguchi system", Expert Systems with Applications, vol. 40, no. 15, pp. 5787-5795.

Kacprzynski, G.J., Roemer, M.J. , Modgil, G. , Palladino, A. and Maynard, K. (2002), "Enhancement of Physics-of-Failure Prognostics Models with System Level Features", in: Aerospace Conference Proceedings, 2002. IEEE, vol.6, pp.6-2919- 6-2925.

Kano, M., Hasebe, S., Hashimoto, I. Strauss, R., Bakshi, B. and Ohno, H. (2000), "Contribution Plots for Fault Identification Based on the Dissimilarity of Process Data", Proc. AIChE Annual Meeting, Los Angeles, CA, USA.

Karabadji, N. E. I., Seridi, H., Khelf, I., Azizi, N. and Boulkroune, R. (2014), "Improved decision tree construction based on attribute selection and data sampling for fault diagnosis in rotating machines", Engineering Applications of Artificial Intelligence, vol. 35, no. 0, pp. 71-83.

Kasabov, N. (2007), "Evolving Connectionist Systems: The Knowledge Engineering Approach", Springer-Verlag New York, Inc.

Kasabov, N. and Song, Q. (2002), "DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and its Application for Time Series Prediction", IEEE Transactions on Fuzzy Systems, vol. 10, no.2, pp.144-154

Kennedy, J. and Eberhart, R. (1995). "Particle Swarm Optimization". Proceedings of IEEE International Conference on Neural Networks IV. pp. 1942–1948.

Kentzoglanakis, K. and Poole, M. (2009), “Particle swarm optimization with an oscillating Inertia Weight”, Proceedings of the 11th Annual conference on Genetic and evolutionary computation, pp. 1749–1750.

Kim, Y.J., Wood, W.H. and Agogino, A.M. (1992), "Signal Validation for Expert System Development", Proceedings of the 2nd International Forum on Expert Systems and Computer Simulations in Energy Engineering, pp. 951-956.

Kohavi, R. and John, G. (1996), "Wrappers for feature subset selection. Artificial Intelligence, special issue on relevance", vol.97, no.1-2, pp.273–324.

Kohavi. R. (1995), "Wrappers for Performance Enhancement and Oblivious Decision Graphs". PhD thesis, Stanford University.

Korbicz, J., Koscielny,J.M., Kowalczuk, Z. and Cholewa, W. (2004), "Fault Diagnosis: Models, Artificial Intelligence, Applications", Springer.

Kulla, J. (2010), "Sensor validation using minimum mean square error estimation", Mechanical Systems and Signal Processing, vol. 24, pp. 1444-1457.

Kumar, A. , Nayak, A. and Patnaik, A. R. (2007), "A combined Model Approach for Prognosis and Life Cycle Management of Aeroengine Components", in: Proceedings of ASME Turbo Expo 2007. Montreal, Canada,GT 2007 – 27495.

Kumar, S., Roy, N. and Ganguli, R. (2007), "Monitoring low cycle fatigue damage in turbine blade using vibration characteristics", Mechanical Systems and Signal Processing, vol. 21, no.1, pp. 480-501.

Lee B.I., Kesler M.G. (1975), "A Generalized Thermodynamic Correlation Based on Three-Parameter Corresponding States", AIChE J., vol.21, no.3, pp.510-527.

Lee, S. C. (1994), "Sensor Value Validation Based on Systematic Exploration of the Sensor Redundancy for Fault Diagnosis KBS", IEEE Transactions on Systems, Man and Cybernetics, vol. 24, no. 4.

Lei, Y., Lin, J., He, Z. and Zuo, M. J. (2013), "A review on empirical mode decomposition in fault diagnosis of rotating machinery", Mechanical Systems and Signal Processing, vol. 35, no. 1–2, pp. 108-126.

Li, B., Zhang, P., Tian, H., Mi, S., Liu, D. and Ren, G. (2011), "A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox", Expert Systems with Applications, vol. 38, no. 8, pp. 10000-10009.

Li, C.J. and Lee, H. (2005), "Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics", Mechanical Systems and Signal Processing, vol.19, pp. 836-846.

Li, H.R. and Gao, Y.L. (2009), “Particle Swarm Optimization Algorithm with Exponent Decreasing Inertia Weight and Stochastic Mutation”, In —2009 Second International Conference on Information and Computing Science, pp. 66–69.

Li, N., Zhou, R., Hu, Q. and Liu, X. (2012), "Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine", Mechanical Systems and Signal Processing, vol. 28, no. 0, pp. 608-621.

Li, P.L. Zhang, S.B. Liang, and Ren, G.Q. (2008), "Feature extraction and selection for fault diagnosis of gear using wavelet entropy and mutual information", The 9th International Conference on Signal Processing, no.1-5, pp. 2843-2847.

Li, X., Zheng, A., Zhang, X., Li, C. and Zhang, L. (2013), "Rolling element bearing fault detection using support vector machine with improved ant colony optimization", Measurement, vol. 46, no. 8, pp. 2726-2734.

Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S. and Liang, S. (1999), "Adaptive Prognostics for Rolling Element Bearing Condition", Mechanical Systems and Signal Processing, vol. 13, no.1, pp.103-113.

Li, Y., Kurfess, T.R. and Liang, S.Y. (2000), "Stochastic Prognostics for Rolling Element Bearings", Mechanical Systems and Signal Processing, vol.14, no.5, pp.747-762.

Li, Z., He, Y., Chu, F., Han, J. and Hao, W. (2006), "Fault recognition method for speed-up and speed-down process of rotating machinery based on independent component analysis and Factorial Hidden Markov Model", Journal of Sound and Vibration, vol. 291, no. 1–2, pp. 60-71.

Lin, C.-J., Hong, S.-J. and Lee, C.-Y. (2006), "The Design of Neuro-Fuzzy Networks Using Particle Swarm Optimization and Recursive Singular Value Decomposition," Neurocomputing, vol. 71, no.1-3, pp. 297-310.

Lin, T. C. and Lee, C. S. (1991), “Neural Network Based Fuzzy Logic Control and Decision System”, IEEE Transactions on Computers, vol. 40, no. 12, pp. 1320-1336.

Literature review of data validation methods. D 3.3.1

Ioannides, E. and Harris, T. A. (1985), "A new Fatigue Life Model for Rolling Bearings", Journal of Tribology, vol.107, pp.367-378.

Luo, J. , Namburu, M. , Pattipati, K. , Liu Qiao, Kawamoto, M. and Chigusa, S. (2003), "Model-based Prognostic Techniques", IEEE Systems Readiness Technology Conference, pp. 330- 340.

Luo, J., Pattipati, K.R. , Liu Qiao and Chigusa, S. (2008), "Model-Based Prognostics Techniques Applied to Suspension System", in: This paper appears in: Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions, vol. 38, no.5, pp. 1156-1168.

Luo, J., Bixby, A., Pattipati, K., Liu Qiao, Kawamoto, M. and Chigusa, S. (2003), "An Interacting Multiple Model Approach to Model-based Prognostics", in: Systems, Man and Cybernetics, 2003. IEEE International Conference, CT, USA, vol.1, pp.189- 194.

Ma, J. and Li, C.J. (1995), "Detection of Localised Defects In Rolling Element Bearings Via Composite Hypothesis test", Mechanical Systems and Signal Processing, vol.9, no.1, pp. 63-75.

Mackey, M.C. and Glass, L. (1977), Oscillation and chaos in physiological control systems, Science, vol.197, pp. 287-289.

Mail, R.F., Rahman, T.A., Hashim, S.Z.M. and Ngah, R. (2007), “New Particle Swarm Optimizer with Sigmoid Increasing Inertia Weight”, International Journal of Computer Science and Security (IJCSS), vol. 1, no. 2, pp. 35.

Malhi, A. and Gao, R. X. (2004), "PCA-Based Feature Selection Scheme for Machine Defect Classification", IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, vol. 53, no.6.

Mamdani, E. H. and Assilian, S. (1975), “An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller”. International Journal of Man-Machine Studies, vol. 7, no. 1, pp.1-13.

Mechefske, C. K. and Mathew, J. (1992), "Fault detection and diagnosis in low speed rolling element bearings ", Mechanical Systems and Signal Processing, vol. 6, no. 4, pp. 259-273.

Meher-Homji, B. C. (1989). “Fouling of Axial Flow Compressors- Causes, Effects, Detection, and Control”, Proceedings, 18th Turbomachinery Symposium, pp.55-76.

Mendel, J. (2001), “Uncertain rule-based fuzzy inference systems: Introduction and new directions”, Prentice-Hall.

Mentzer, J.T. (1988), "Forecasting with Adaptive Extended Exponential Smoothing", Academy of Marketing Science, vol.16 , no. 3-4, pp. 062-070.

Millioz, F., Martin, N., Leonard, F. and Vieira, M. (2011), "Circularity of the STFT and spectral kurtosis for time-frequency segmentation in Gaussian environment", IEEE Trans. Signal Process. Vol.59, no. 2,pp.515-524.

Mori, K., Kasashima, N., Yoshioka, T. and Ueno, Y. (1996), "Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals", Wear, vol. 195, no. 1–2, pp. 162-168.

Mostafa M. Awad (2011). Fouling of Heat Transfer Surfaces, Heat Transfer - Theoretical Analysis, Experimental Investigations and Industrial Systems, Prof. Aziz Belmiloudi (Ed.), ISBN: 978-953-307-226-5.

Moubray, J. (1997), "Reliability-Centred Maintenance", (2nd ed), Industrial Press Inc. New York.

Muralidharan, V. and Sugumaran, V. (2013), "Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump", Measurement, vol. 46, no. 1, pp. 353-359.

Nikabadi, A. and Ebadzadeh, M. (2008), “Particle swarm optimization algorithms with adaptive Inertia Weight : A survey of the state of the art and a Novel method”, IEEE journal of evolutionary computation.

Orsagh R., Roemer, M., Sheldon, J. and Klenke, C.J. (2004) "A Comprehensive Prognostics Approach for Predicting Gas Turbine Engine Bearing Life", in: Proceedings of IGTI Turboexpo 2004. Vienna. GT2004-53965.

Orsagh, R.F. ,Sheldon, J. and Klenke, C.J.(2003), "Prognostics/Diagnostics for Gas Turbine Engine Bearings", in: Aerospace Conference, 2003. Proceedings. 2003 IEEE, vol.7, pp. 3095- 3103.

Pan, Y., Chen, J. and Li, X. (2010), "Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means", Mechanical Systems and Signal Processing, vol. 24, no. 2, pp. 559-566.

Platt, J. (1991), "A Resource Allocating Network for Function Interpolation", Neural Comp., vol.3, pp. 213-225.

Pouliezos A. D. and Stavrakankis, G. S. (1994), "Real Time Fault Monitoring of Industrial Processes", Kulwer Academic Publishers.

Poyhonen, S., Jover, P., and Hyotyniemi, H. (2004), "Signal Processing of vibrations for condition monitoring of an induction motor", Proceedings of the first international symposium on control, communications and signal processing, pp. 499-502.

Pratama, M., Er, M. J., Li, X., Oentaryo, R. J., Lughofer, E. and Arifin, I. (2013), "Data driven modeling based on dynamic parsimonious fuzzy neural network", Neurocomputing, vol. 110, no. 0, pp. 18-28.

Qin, S.J. and Li, W. (1999), "Detection, identification, and reconstruction of faulty sensors with maximised sensitivity", AIChE Journal. Vol.45, no. 9, pp.1963-1976.

Rafiee, J., Arvani, F., Harifi, A. and Sadeghi, M. H. (2007), "Intelligent condition monitoring of a gearbox using artificial neural network", Mechanical Systems and Signal Processing, vol. 21, no. 4, pp. 1746-1754.

Rai, P. and Singh, S. (2010), “A Survey of Clustering Techniques”, International Journal of Computer Applications, vol. 7, no. 12, pp.1–5. Published By Foundation of Computer Science

Rajput ,R. K. (2009), Engineering Thermodynamics, Infinity Science Series / Engineering series (3 ed.), Jones & Bartlett Learning, pp. 77.

Rosenblatt, F. (1957), “The Perceptron--a perceiving and recognizing automaton”. Report 85-460-1, Cornell Aeronautical Laboratory.

Ross, G.J. , Adams, N.M., Tasoulis, D.K. and Hand, D.J. (2012), "Exponentially weighted moving average charts for detecting concept drift", Pattern Recognition Letters, vol.33, no. 2, pp. 191–198

Ross, T.J. (2004), “Fuzzy logic with engineering applications”, John Wiley & Sons, Ltd.

Rousseeuw, P.J. and Zomeren, B.C.V. (1990), "Unmasking multivariate outliers and leverage points", Journal of the American Statistical Association, vol.85, no.411, pp.633- 639.

Sakthivel, N. R., Sugumaran, V. and Babudevasenapati, S. (2010), "Vibration based fault diagnosis of monoblock centrifugal pump using decision tree", Expert Systems with Applications, vol. 37, no. 6, pp. 4040-4049.

Sakthivel, N. R., Sugumaran, V. and Nair, B. B. (2010), "Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump", Mechanical Systems and Signal Processing, vol. 24, no. 6, pp. 1887-1906.

Samanta, B. and Nataraj, C. (2008), "Prognostics of machine condition using soft computing", Robotics and Computer-Integrated Manufacturing, vol. 24, no. 6, pp. 816-823.

Samuel, P. D. and Pines, D. J. (2005), "A review of vibration-based techniques for helicopter transmission diagnostics", Journal of Sound and Vibration, vol. 282, no. 1–2, pp. 475-508.

Saravanan, N. and Ramachandran, K.I. (2009) , "Fault diagnosis of spur bevel gear box using discrete wavelet features and decision tree classification", Expert Systems with Applications, vol.36, pp.9564-9573.

Saravanan, N., Cholairajan, S., and Ramachandran, K.I. (2009), "Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique", Expert Systems with Applications, vol.36, pp. 3119–3135.

Sathiyasekar, K., Thyagarajah, K. and Krishnan, A. (2011), "Neuro fuzzy based predict the insulation quality of high voltage rotating machine", Expert Systems with Applications, vol. 38, no. 1, pp. 1066-1072.

Sekhar, A.S. (2004), "Model based identification of two cracks in a rotor system", Mechanical Systems and Signal Processing, vol.18, pp.977-983.

Shao, Y. and Mechefske, C. K. (2009), "Gearbox vibration monitoring using extended Kalman filters and hypothesis tests", Journal of Sound and Vibration, vol. 325, no.3, pp. 629-648.

Shi, Y. and Eberhart, R. (2002), “A modified particle swarm optimizer”, In Evolutionary Computation Proceedings, 1998. IEEE World Congress on

Computational Intelligence, The 1998 IEEE International Conference on, pp. 69–73.

Shieh, A.D. and Hung, Y.S. (2009), "Detecting outlier samples in microarray data" , Statistical Applications in Genetics and Molecular Biology, vol. 8, no. 1.

Shmueli, G. 2011. Practical Time Series Forecasting: A Hands-On Guide. 2nd ed. s.l.:Axelrod Schnall.

Snider, S., (2006), “TA 2002 Ethylene Plant Cracker Gas Compressor Fouling”, AIChE Spring National Meeting, EPC Conference, Texas, Houston.

Song, Q. and Kasabov, K. (2001), "ECM - A Novel On-line, Evolving Clustering Method and Its Applications", In M. Posner (Ed.), Foundations of cognitive science, The MIT press, pp. 631-682.

Song, Q., Kasabov, N. (2001), “ECM – A Novel On-line Evolving Clustering Method and Its Applications” In: Posner, M.I. (ed.) Foundations of cognitive science, pp. 631–682.

Song, T.W. (2005). “Predictions of The Performance Degradation of Industrial Gas Turbines Due to Compressor Fouling”, 16th International Symposium on Transport Phenomena, Prague.

Specht, D.F. (1991), "A general Regression Neural Network", IEEE Transactions on Neural Networks, vol.2, no.6, pp.568-576.

Stack, J.R., Harley, R.G. and Habetler, T.G. (2002) , "An amplitude modulation detector for fault diagnosis in rolling element bearings", Shock and Vibration, vol.9, pp.225-234.

Staszewski, W., Ruotolo, R. and Storer, D. (1999), "Fault Detection in Ball Bearings Using Wavelet Variance", Proc. Of Int. Modal Analysis Conference-IMAC, vol. 2, pp. 1335-1339

Su, Z., Tang, B., Ma, J. and Deng, L. (2014), "Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier", Measurement, vol. 48, no. 0, pp. 136-148.

Sugeno, M., (1985), "Industrial applications of fuzzy control", Elsevier Science Pub. Co.

Sugumaran, V. and Ramachandran, K. I. (2011), "Effect of number of features on classification of roller bearing faults using SVM and PSVM", Expert Systems with Applications, vol. 38, no. 4, pp. 4088-4096.

Sun, Q., Chen, P., Zhang, D., and Xi, F. (2004), "Pattern recognition for automatic machinery fault diagnosis", Journal of Vibration and Acoustics, vol. 126, pp. 307-316.

Sun, W., Chen, J. and Li, J. (2007), "Decision tree and PCA-based fault diagnosis of rotating machinery", Mechanical Systems and Signal Processing, vol. 21, no. 3, pp. 1300-1317.

Swain, M., Dash, S.K., Dash, S. and Mohapatra, A. (2012). "An approach for iris plant classification using neural networks". International Journal on Soft Computing (IJSC), vol.3, no.1.

Takagi, T. and Sugeno, M. (1985), "Fuzzy identication of systems and its applications to modelling and control", IEEE transactions on systems, man, and cybernetics , vol. 15,no. 1, pp. 116-132.

Theophano , M.(2010), “Temporal Data Mining”, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series.

Thiry, C., Yan, A. M. and Golinval, J. C. (2004), "Damage Detection in Rotating Machinery Using Statistical Methods: PCA Analysis and Autocorrelation Matrix", Surveillance 5 CETIM Senlis.

Time Series." Merriam-Webster.com. Merriam-Webster, n.d. Web. 17 July 2014.

Timofeev, R. (2004), "Classification and Regression Trees (CART) Theory and Applications", (Unpublished Master thesis), Humboldt University, Berlin.

Tran, V. T., AlThobiani, F. and Ball, A. (2014), "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks", Expert Systems with Applications, vol. 41, no. 9, pp. 4113-4122.

Tran, V. T., Thom Pham, H., Yang, B. and Tien Nguyen, T. (2012), "Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine", Mechanical Systems and Signal Processing, vol. 32, no. 0, pp. 320-330.

Tran, V. T., Yang, B., Gu, F. and Ball, A. (2013), "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis", Mechanical Systems and Signal Processing, vol. 38, no. 2, pp. 601-614.

Tran, V.T., Yang, B.-S. and Tan, A.C. C. (2009), "Multistep ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems", Expert Systems with Applications, vol.36, no.5, pp.9378-9387.

Tse, P.W. and Atherton, D.P.(1999), "Prediction of Machine Deterioration Using Vibration Based Fault Trends and Recurrent Neural Networks", Journal of Vibration and Acoustics, vol.121, no.3, pp.355-362.

Tukey,J. W. (1977),Exploratory Data Analysis, Addison Wesley.

Turner, B., “On-Stream Cleaning Of Turbomachinery”, Proceedings, 2nd Turbomachinery Symposium, Gas Turbine Laboratories, Texas A&M University.

Usynin, A. (2006). "Model-Fitting Approaches to Reliability Assessment and Prognostic Problems", Journal of Pattern Recognition Research, vol.1, pp. 32-36.

Vachtsevanos, G. and Wang, P. (2001), "Fault Prognosis Using Dynamic Wavelet Neural Networks", in: AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference, Valley Forge, PA, USA, pp. 857-870.

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A. and Wu, B. (2007), Intelligent Fault Diagnosis and Prognosis for Engineering Systems, John Wiley & Sons, Inc.

Volpi, S., Cococcioni, M., Lazzerini, B., and Stefanescu, D. (2010), "Rolling element bearing diagnosis using convex hull", In International joint conference on neural networks,pp. 1–8.

Wang X.Y., Makis, V. and Yang, M. (2010), "A wavelet approach to fault diagnosis of a gearbox under varying load conditions", Journal of Sound and Vibration, vol.329, pp.1570-1585.

Wang, C., Zhou, J., Qin, H., Li, C. and Zhang, Y. (2011), "Fault diagnosis based on pulse coupled neural network and probability neural network", Expert Systems with Applications, vol. 38, no. 11, pp. 14307-14313.

Wang, J. and Hu, H. (2006), "Vibration-based fault diagnosis of pump using fuzzy technique", Measurement, vol. 39, no. 2, pp. 176-185.

Wang, L.-X. (1994), “Adaptive fuzzy systems and control: design and stability Analysis”, Prentice Hall.

Wang, W. and Kanneg, D. (2009), "An integrated classifier for gear system monitoring", Mechanical Systems and Signal Processing, vol. 23, no. 4, pp. 1298-1312.

Wang, W.Q., Golnaraghi, M.F. and Ismail,F. (2004), "Prognosis of machine health condition using neuro-fuzzy systems", Mechanical Systems and Signal Processing, vol.18 , pp.813-831.

Ward ,J.H. (1963), "Hierarchical grouping to optimize an objective function", Journal of American Statistical Association, vol. 58, pp. 236–244.

Wegerich, S. (2005), "Similarity Based Modeling of Vibration Features for Fault Detection and Identification", Sensor Review, vol. 25, no. 2.

Wegerich, S. (2006), "Condition Based Monitoring using Nonparametric Similarity Based Modeling", Proceedings of the 3rd annual meeting of Japan Society of Maintenology, pp. 308-313.

Western Electric Company (1956), "Statistical Quality Control handbook. (1 ed.) ", Indianapolis, Indiana: Western Electric Co., p. v.

Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T. and Vapnik, V. (2001), "Feature selection for SVMs", Advances in Neural Information Processing Systems, vol.13.

Wu, J. and Liu, C. (2009), "An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network", Expert Systems with Applications, vol. 36, no. 3, Part 1, pp. 4278-4286.

Wu, J., Huang, C., Chang, Y. and Shiao, Y. (2010), "Fault diagnosis for internal combustion engines using intake manifold pressure and artificial neural network", Expert Systems with Applications, vol. 37, no. 2, pp. 949-958.

Wu, J.D. and Hsu, C.C. (2009), "Fault gear identification using vibration signal with discrete wavelet transform technique and fuzzy-logic inference", Expert systems with applications, vol.36, pp.3785-3794.

Wu, J.D. and Kuo, J.M (2009), "An automotive generator fault diagnosis system using discrete wavelet transform and artificial neural network", Expert Systems with Applications, vol. 36, no.6, pp.9776-9783.

Wu, J.D. and Liu, C.H. (2008), "Investigation of engine fault diagnosis using discrete wavelet transform and neural network", Expert Systems with Applications, vol.35, pp. 1200-1213.

Xin, J., Chen, G. and Hai, Y. (2009), “A Particle Swarm Optimizer with Multistage Linearly-Decreasing Inertia Weight”, Computational Sciences and Optimization, 2009. CSO 2009. International Joint Conference on, vol. 1, pp. 505–508. IEEE.

Xu, Z., Xuan, J., Shi, T., Wu, B. and Hu, Y. (2009), "A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique", Expert Systems with Applications, vol. 36, no. 9, pp. 11801-11807.

Yan, J., Guo, C. and Wang, X. (2011), "A dynamic multi-scale Markov model based methodology for remaining life prediction", Mechanical Systems and Signal Processing, vol. 25, no. 4, pp. 1364-1376.

Yang, G., Jian, Q. and Zhou, H. (2014), "Research on Bearing Fault Diagnosis Using APSO-SVM Method", Sensors & Transducers, vol. 175, no. 7, pp. 207-213.

Yang, J., Xiong X., and Xiong, S., (1999), "Damage Detection of Roller Bearing using Wavelet Transforms", Proc. Of Int. Modal Analysis Conference-IMAC, vol. 2, pp. 1980-1983.

Yayar, R., Hekim, M., Yilmaz , V. and Bakirci,F. (2011), “A comparison of ANFIS and ARIMA techniques in the forecasting of electric energy consumption of Tokat province in Turkey”. Journal of Economic and Social Studies, vol.1, no.2, pp.87-112.

Ye, J. (2009), "Fault diagnosis of turbine based on fuzzy cross entropy of vague sets", Expert Systems with Applications, vol. 36, no. 4, pp. 8103-8106.

Yiakopoulos, C.T., Gryllias , K.C. and Antoniadis, I.A. (2011),"Rolling element bearing fault detection in industrial environments based on a K -means clustering approach", Expert Systems with Applications, vol.38,no.3, pp.2888–2911.

Younus, A. M. D. and Yang, B. (2012), "Intelligent fault diagnosis of rotating machinery using infrared thermal image", Expert Systems with Applications, vol. 39, no. 2, pp. 2082-2091.

Yu, D.J., Cheng, J.S. and Yang, Y. (2005), “Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings,” Mechanical Systems Signal Processing, vol. 19, no. 2,pp.259-270

Yuan, S. and Chu, F. (2006), "Support vector machines-based fault diagnosis for turbo-pump rotor", Mechanical Systems and Signal Processing, vol. 20, no. 4, pp. 939-952.

Zadeh, L. (1965). "Fuzzy sets", Information and Control, vol. 8, no. 3, pp.338–353.

Zarei, J. (2012), "Induction motors bearing fault detection using pattern recognition techniques", Expert Systems with Applications, vol. 39, no. 1, pp. 68-73.

Zhang, L., B. Jack, L. and Nandi, A. K. (2005), "Fault detection using genetic programming", Mechanical Systems and Signal Processing, vol. 19, no. 2, pp. 271-289.

Zhao, J. and Bose, B.K. (2002), “Evaluation of membership functions for fuzzy logic controlled induction motor drive”, Industrial Electronics Society, IEEE, 28th Annual Conference, vol.1, pp.229-234.

Zio, E. and Maio, F.D. (2010), "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system", Reliability Engineering and System Safety, vol.95, no.1, pp.49-57.

APPENDICES

Appendix A Classification Dataset

A.1 Dissolved Gas Analysis (DGA) Dataset

C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	Output
1.000	0.046	0.032	1
1.167	0.107	0.100	1
2.000	0.131	0.009	1
0.000	0.019	0.034	1
1.000	0.027	0.001	1
0.000	0.685	0.013	1
2.154	0.256	1.182	2
3.360	0.328	4.879	2
18.538	0.171	8.667	2
4.574	0.221	10.024	2
2.970	0.133	8.630	2
2.882	0.133	8.462	2
1.000	0.167	1.000	2
6.111	0.150	11.392	2
5.779	0.535	12.264	2
5.294	0.125	0.300	2
1.167	0.105	1.000	2
3.000	0.156	6.625	2
6.020	0.163	39.200	2
1.866	0.162	2.629	2
2.742	0.134	9.889	2
5.000	0.208	8.000	2

C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	Output
2.979	0.358	17.843	2
1.602	0.482	2.126	2
1.842	0.324	1.047	2
2.758	0.008	3.300	2
7.429	0.250	3.500	2
8.077	0.150	30.871	2
1.111	0.421	13.500	2
2.490	0.202	16.000	3
1.833	0.205	8.500	3
1.850	0.391	14.399	3
1.562	0.239	9.563	3
1.455	0.201	12.474	3
1.348	0.524	4.763	3
1.139	0.303	12.625	3
0.896	0.453	21.274	3
1.028	0.707	10.171	3
0.668	1.625	11.765	3
1.737	0.223	10.429	3
2.230	0.299	19.063	3
2.000	0.231	3.000	3
1.962	0.489	7.571	3
4.633	0.173	20.750	3
1.143	0.618	12.250	3
1.728	0.252	8.789	3
0.545	0.867	6.111	3
1.262	0.462	18.480	3
0.829	0.827	9.543	3

C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	Output
1.857	0.200	2.000	3
0.940	0.651	3.129	3
1.000	0.083	10.500	3
1.032	0.311	3.875	3
1.412	0.350	7.727	3
1.917	0.254	17.647	3
1.509	0.595	12.927	3
1.029	1.000	14.957	3
0.950	1.717	10.000	3
1.246	0.742	11.296	3
2.308	0.200	11.304	3
0.818	0.263	14.107	3
1.966	0.650	15.676	3
0.660	0.279	7.880	3
1.133	0.279	10.943	3
1.786	0.564	1.556	3
1.164	0.457	21.953	3
1.072	0.238	13.185	3
1.275	0.490	7.278	3
1.042	0.218	17.333	3
0.708	0.807	4.283	3
0.860	0.716	5.135	3
1.100	0.240	12.048	3
1.419	0.673	21.246	3
1.308	0.468	15.287	3
0.006	2.717	2.673	4
0.005	2.301	4.660	4

C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	Output
0.035	1.694	1.004	4
0.250	27.000	0.082	4
0.001	1.739	3.076	4
0.000	12.708	0.345	4
0.000	1.500	1.000	4
0.000	0.909	3.500	4
0.189	0.648	1.526	4
0.004	4.167	1.209	4
0.037	2.470	1.601	4
0.143	3.143	0.056	4
0.000	0.073	0.150	4
0.000	2.240	3.799	4
0.010	0.010	8.571	4
0.000	7.280	1.326	5
0.042	1.565	12.643	5
0.013	1.455	9.095	5
0.031	3.331	6.054	5
0.000	4.200	2.818	5
0.020	1.942	68.333	5
0.183	0.147	6.667	5
0.029	2.350	3.905	5
0.016	2.000	6.091	5
0.010	4.345	3.550	5
0.034	1.768	6.679	5
0.204	1.097	9.649	5
0.151	1.850	5.665	5
0.060	8.000	12.500	5

C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	Output
0.021	2.333	6.071	5
0.009	1.336	6.529	5
0.000	1.000	0.287	6
0.100	2.000	1.000	6
0.060	1.190	2.338	6
0.300	0.500	0.769	6
0.059	0.700	2.429	6
0.015	0.250	4.000	6
0.150	0.000	0.500	6
0.133	0.800	1.500	6
0.000	0.015	4.000	6
0.067	2.947	0.600	6
0.050	0.667	1.200	6
0.337	0.940	3.192	6
0.136	1.682	1.222	6
2.317	0.766	0.537	6
0.600	0.600	1.667	6
0.150	0.250	4.000	6
0.688	1.672	0.407	6
0.750	0.000	1.000	6
0.500	0.500	1.000	6
0.720	0.760	1.389	6
0.727	0.467	2.200	6
1.064	0.868	3.425	6
0.333	0.194	0.375	6
0.000	0.318	1.250	6
0.500	0.200	1.000	6

C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	Output
0.333	1.833	0.429	6
0.025	0.800	0.308	6
0.500	1.500	0.160	6
0.500	0.100	0.160	6

Appendix B Matlab Sample Codes

B.1 Data Clustering

```
n0=size(trndata,1);

[nosamples,nofeatures]=size(trndata);

clusterIndex=size(Cluster,1);

MAT1(:,1:nofeatures)=Cluster;

MAT1(:,nofeatures+1)=Radii;

for i=1:n0

nocls=size(MAT1,1);

for j=1:nocls

MAT1(j,nofeatures+2)=sqrt(((trndata(i,:)-CAT(j,1:nofeatures))*(trndata(i,:)-
MAT1(j,1:nofeatures))')/nofeatures);

end

[minDist,minDistIndex]=min(MAT1(:,nofeatures+2));

if minDist < MAT1(minDistIndex,nofeatures+1)

disp('Point belongs to cluster:')

num(i)=minDistIndex;

class(i)=Output(minDistIndex);

else

for j=1:nocls

MAT1(j,nofeatures+3) = MAT1(j,nofeatures+1) + MAT1(j,nofeatures+2);

end

[minVal,minValIndex] = min(M1AT(:,nofeatures+3));

disp('minVal, minValIndex:')

num(i)=minValIndex;

 if minVal>(2*Dthr)

disp('lowest Value is greater than 2*threhold; add a cluster')

clusterIndex = clusterIndex +1;

MAT1(clusterIndex,1:nofeatures) =trndata(i,:);
```

```
MAT1(clusterIndex,nofeatures+1) = 0;
num(i)=length(MAT1(:,nofeatures+1));
class(i)=1.5;
Output(clusterIndex)=1.5;
else
disp('broaden radius and move cluster center')
MAT1(minValIndex,nofeatures+1) = (minVal/2); % new radius
MAT1(minValIndex,1:nofeatures)=trndata(i,:)*((minDist-
MAT1(minValIndex,nofeatures+1))/(minDist))+
MAT1(minValIndex,1:nofeatures)*((
MAT1(minValIndex,nofeatures+1))/(minDist));
num(i)=minValIndex;
class(i)=Output(minValIndex);
end
end
end
Cci0=MAT1(:,1:nofeatures);
Rui0=MAT1(:,nofeatures+1);
```

B.2 Dynamic Evolving Neurofuzzy Inference System

```
m=size(Ccj,1);
n=size(Ccj,2);
nosamples=size(data,1);
data0=traindatain(1:nosamples,:);
 for i=1:nosamples
 for j=1:m
 D(i,j)=sqrt(((data0(i,:)-Ccj(j,:))*(data0(i,:)-Ccj(j,:))')/n);
 end
 end
 for j=1:m
```

```
 for k=1:n
 c(j,k)=Ccj(j,k);

end
end
Bw=zeros(size(Ccj,2)+1,size(Ccj,1));
Pwold=[];
for i=1:m
[G Index]=sortrows(D,i);
H(:,1)=Index;
H(:,2)=G(:,i);
MIndex=H(1:n*pq,:);
data1=traindatain (Index(1:n*pq),:);
for k=1:size(MIndex,1)
A(k,2:n+1)=data1(k,:);
A(k,1)=1;
end
for t=1:size(MIndex,1)
W(t,t)=1-MIndex(t,2);
end
y=traindataout(Index(1:n*pq));
 Pw=inv(A'*W*A)
 bw(:,i)=Pw*A'*W*y;
 Pw1=[Pwold Pw];
 Pwold=Pw1;
 end
 temp=zeros(size(Ccj,1),size(data0,2));
 for i=1:size(data0,1)
 for j=1:size(Ccj,1)
```

```
for k=1:size(data0,2)
x=data0(i,k);
temp(j,k)=exp(-1*(x-c(j,k))^2/(2*sigma^2));
end
end
mu=prod(temp,2);
ak(1)=1;
ak(2:n+1)=data0(i,:);
fi=ak*bw;
output(i)=(sum(fi*mu))/sum(mu);
end
output=output';
```